全国第六届研究生数学建模竞赛

题 目　　我国城镇就业人数的数学模型

摘　　要：

本文主要讨论的是金融危机大背景下我国的就业问题。在充分理解问题的基础上，利用 Granger 因果检验、人工神经网络、计量经济学等知识对各问进行分析，建立模型求解得到所需结果，并对结果的合理性做出验证。

对于第一问，利用 Granger 因果检验法，并从 15 个粗略指标中成功提炼出 6 个最能影响就业的指标：人均 GDP、第三产业产值、社会零售商品总额、职工平均工资、政府财政支出以及总人口数。为后几问打下基础。

对于第二问，寻求城镇就业人数与上述 6 个指标之间的定量关系。利用人工神经网络良好的函数逼近性，逐一试探径向基函数网络，BP 网络和改进的 BP 网络，发现改进的 BP 网络是最适合的方法。

对于第三问，认为是全文的重点。建立 4 个数学模型：经济社会指标和就业的可加性分解模型，行业和就业的 VEC 经济模型，地区和就业的主成分分析和层次聚类分析，黑龙江省大学生就业率最小二乘拟合模型。

对于第四问，通过国家决策和规划以及多项式拟合得到输入数据，利用第二问建立的因果检验与神经网络整合模型，进行仿真，得出的结果即为 2009 年和 2010 年就业形势预测。

对于第五问，纵观前面的建模和仿真过程，有针对性地为金融危机大背景下的国家宏观调控、企业决策、个人发展等提出较为合理的建议。

本文的创新点：通过有限次试探，寻求最合适选择；因果检验与神经网络相结合；将计量经济学理论用来解决实际问题等。

关键词：就业人数　Granger 因果检验　人工神经网络　可加性分解模型　VEC 模型　最小二乘拟合　协整检验

参赛队号 1021308

队员姓名 侯月明　夏晴　贺智

参赛密码 _____

（由组委会填写）

1 问题重述

就业问题是世界各国政府所面临的重大社会问题，更是中国社会主义和谐社会建设的关键所在。失业、经济增长和通货膨胀为宏观经济中特别重要的三个指标。按照已有研究，就业可以定义为三个月内有稳定的收入或与用人单位有劳动聘用关系。失业的统计方法各国差异较大,我国采用城镇登记失业率，是指城镇登记失业人数同城镇从业人数与城镇登记失业人数之和的比。其中，城镇登记失业人员是指有非农业户口，在一定的劳动年龄内（16 岁以上及男 50 岁以下、女 45 岁以下），有劳动能力，无业而要求就业，并在当地就业服务机构进行求职登记的人员。但由于统计口径不同，存在一定的差异，有些历史数据也较难获得。

从经济学的角度，影响就业（或者失业）的因素很多。从宏观层面上，消费、投资、政府购买和进出口都是重要的因素；而从中观层面，不同地区、不同产业也会表现出不同的特征。当然，中央政府调整宏观经济政策（包括财政政策和货币政策），以及对不同地区和不同产业实行不同的扶持政策都会对就业产生巨大的影响。

就我国的现实情况，2008 年我国经济社会经受了历史罕见的考验，GDP 依然保持 9%以上平稳较快增长，城镇新增就业 1113 万人，城镇登记失业率为 4.2%。2009 年我国就业面临更大的挑战，一是国际金融危机导致国际市场需求难以在短期内复苏；二是今年我国经济增速下滑；三是国内消费需求乏力；四是一些行业产能过剩与市场预期不确定导致企业投资不足，所以就业形势十分严峻。

为此，中央政府从 08 年 10 月开始实施了 40000 亿元的投资计划，确定了十大产业振兴计划，采取扩大国内消费需求的措施，提高对外开放水平以增加出口。同时，中央财政拟投入 420 亿元资金实施积极的就业政策。09 年我国在就业方面的目标：城镇新增就业 900 万人以上，城镇登记失业率控制在 4.6%以内（以上数据取自温家宝总理的政府工作报告）。

利用近年来我国有关的统计数据并结合一年多来我国国民经济的运行数据（参见下面网站，也可以对比其他国家的统计数据）就我国就业人数或城镇登记失业率研究如下问题。

1．对有关统计数据进行分析，寻找影响就业的主要因素或指标。

2．建立城镇就业人数或城镇登记失业率与上述主要因素或指标之间联系的数学模型。

3．对上述数学模型从包含主要的经济社会指标、分行业、分地区、分就业人群角度，尝试建立比较精确的数学模型。（由于时间限制，建议适度即可）

4．利用所建立的关于城镇就业人数或城镇登记失业率的数学模型，根据国家的有关决策和规划对 2009 年及 2010 年上半年的我国就业前景进行仿真（可以根据模型的需要对未来的情况作适当的假设）。

5．根据所建立的数学模型和仿真结果，对提高我国城镇就业人口数或减少城镇登记失业率提出咨询建议。

2 问题分析

随着经济在全球化的不断发展和国内改革发展进入关键时期，就业无疑是当前社会最大的问题。作为关系发展和改革全局的重大问题，就业已经越来越凸显对社会经济发展的重要性。十七大报告提出了实施扩大就业的发展战略，强调“完善支持自主创业、自谋职业政策，加强就业观念教育，使更多劳动者成为创业者”。虽然近年来中国就业再就业工作取得了显著成效，但劳动力供大于求的趋势长期存在，“十一五”期间城镇就业每年供大于求 1200 万人左右。当前的就业形势依然严峻，突出表现在城镇新增就业同比增速下降，登记失业率上升，高校毕业生就业难问题更为突出等问题。经济结构调整、人口峰值逼近等一系列问题也不断加大就业再就业工作的压力。

问题一的分析：

就业问题是一个极其复杂的问题，很多因素都可能对它产生或大或小的影响。宏观层面上有消费、投资、政府购买和进出口等，中观层面上有地区差异、产业结构的等，另外，中央政府调整宏观经济政策（包括财政政策和货币政策），以及对不同地区和不同产业实行不同的扶持政策都会对就业产生巨大的影响。

为了能尽可能全面地考虑到各种影响就业的因素，我们从中提取出了如下 15 个要素：人均 GDP、城镇就业人口比重、货物进出口总额、全社会固定资产投资额、政府财政支出、商品零售总额、通货膨胀率、第一产业生产总值、第二产业生产总值、第三产业生产总值、消费指数、发电量、钢材产量、能源生产总量、教育经费。并由《中国统计年鉴》等有关资料，得到从 1978 年至 2008 年数据。

由于相关性分析只能判断出趋势容易出现伪回归现象，通过格兰杰因果因果检验判断各因素与就业人数的相关性大小。通过建立滞后自回归模型，寻找出影响就业的 6 个主要因素或指标。这些指标将作为后续问题的自变量，建立数学模型。

问题二的分析：

要利用问题一已经得到的 6 个主要指标，寻求城镇就业人数与这些指标之间的定量关系。对未来就业人数的变化进行准确分析，为政府未来的就业政策、现实人口政策和社会保障政策的科学制定提供依据。

还有一点需要说明的是，本文选取的因变量是城镇就业人数而不是城镇登记就业率。这是因为本文作者认为选择城镇就业人数相比城镇登记就业人数更合理。考虑到失业者不一定都到相关部门进行了登记，即使登记了，也不一定在再次找到工作时取消了登记。由此看来，城镇就业人数的数据统计会比城镇登记就业率准确。

利用人工神经网络（ANN）良好的函数逼近性，对城镇就业人数与上述 6 个主要指标的关系，逐一试探径向基函数神经网络，BP 神经网络，改进的 BP 神经网络，通过比较，选取最适合的神经网络方法，构建因果关系理论与神经网络的整合模型。

问题三的分析：

此问是全文的重点。为了从包含主要的经济社会指标、分行业、分地区、分就业人群的角度，分析城镇就业人数，同时，考虑到问题的复杂性，可以先分别考虑经济社会指标、行业、地区、就业人群 4 个独立方面与城镇就业人数的关系。

经济社会指标作为衡量我国经济社会发展状况的重要指标之一[1]，几乎包含了经济社会的各项指标，本文中重点考虑其中城镇人均 GDP、经济增长率、产业结构变化、进出口总额、中间需求（与消费品的数量有关）、国内需求。

对于行业的理解，看作是产业结构因不同，对城镇就业人数造成影响。而对于产业结构的分类方法，在经济研究和经济管理中，经常使用的分类方法主要有两大领域、两大部类分类法，三次产业分类法。本文中，我们采用三次产业分类法：把产业分为第一产业、第二产业、第三产业[2]。这也可以看成是对行业的粗略划分，有利于简化模型。

而对于 GB/T4754-94 国家行业分类标准中所划分的 16 个行业，将作为探求城镇就业人数与地区关系的影响因素。

对城镇就业人数与地区的关系，重点寻找地区差异对就业造成影响的规律。考虑到比赛时间问题，仅对全国进行就业相关的区域划分。

对就业人群与就业的关系，理解为受教育度不同对就业造成的影响。本文重点考虑与大学生息息相关的就业率问题。

首先，建立主要经济社会指标与就业人数的可加性分解模型。

然后，建立第一、第二、第三产业与就业人数的 VEC 模型。

再次，利用主成分分析法和层次聚类法，对全国进行区域划分。

最后，利用最小二乘法拟合黑龙江省大学生就业率，并对未来两年进行预测。

问题四的分析：

按照题目要求，利用第二问所建立的关于城镇就业人数的数学模型，对 2009 年及 2010 年上半年的我国就业前景进行仿真。其中，对于各指标的数据，对能从国家的有关决策和规划中查到的数据，就假定此数据为真实的，亦即假定国家的有关决策和规划能够达到。对于没有从国家相关政策和规划中查到的数据，可考虑根据历年的数据，通过多项式拟合进行预测。

问题五的分析：

在前四问已经建立模型和仿真（仿真过程充分考虑了国家政策调整所产生的影响）结果的基础之上，进行简单的定性分析，有针对性地为金融危机大背景下的国家宏观调控，企业发展和个人前途等提出较为合理的建议。

3 模型假设

1) 假设 1 仅考虑城镇就业人数，不考虑农村就业人数；
2) 假设 2 劳动年龄（16 岁以上及男 50 岁以下、女 45 岁以下）；
3) 假设 3 统计数据真实可靠；
4) 假设 4 国家相关决策和规划能完成。

4 符号说明

表 1 符号说明

符号	描述
L_t	社会所能容纳的总就业人数
X_t	t 时期的工作产出
$A_1, A_2, \cdots A_p$	参数矩阵
ε_t	扰动向量
Y_t	内生变量组成的向量
P	滞后项数
X, Y, E, M	总产出、最终需求、出口和竞争性进口的 n 维列向量
A	直接消耗系数矩阵

5 模型的建立与求解

5.1 问题一

（1）相关性分析

在研究变量间的关系时，常用相关系数表示变量间相关程度的强弱。典型相关分析是由霍特林提出，其目的是识别并量化两组变量之间的联系，将两组变量相关关系的分析转化为一组变量的线性组合与另一组变量的线性组合之间的相关关系分析[3]。

相关性分析的基本想法是：首先求第一对线性组合，使得它们有最大的简单相关系数（两个变量间的线性相关性用简单相关系数表示）；然后从与第一对线性组合不相关的线性组合对中选出有最大相关系数的一对，成为第二对线性组合，以此类推[4,5]。

本文中，由于要得到城镇就业人数与各因素间的相关性，首先自然而然地想到将每个变量分别与城镇就业人数进行相关性分析，以便通过相关系数的大小来判断变量间的关系。

以人均国内生产总值（GDP）为例，进行相关性检验，采用 SPSS 软件。

首先通过绘制 Q-Q 图的方法判断了两个变量是否符合正态分布。

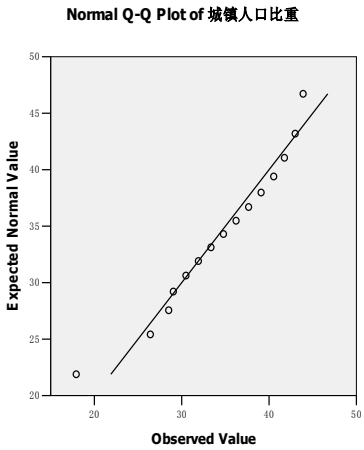


图 1 人口比重 QQ 图

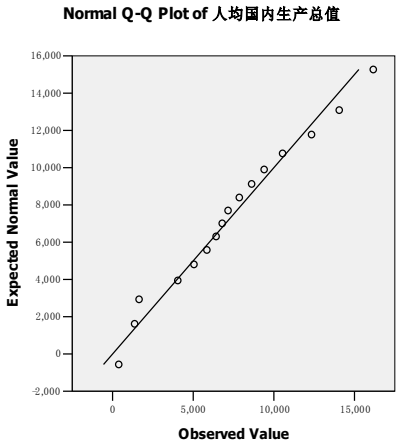


图 2 人均 GDPQQ 图

由图1和图2可知，两个变量均符合正态分布，所以采用Pearson算法进行相关性分析，得到的界面如下图所示。

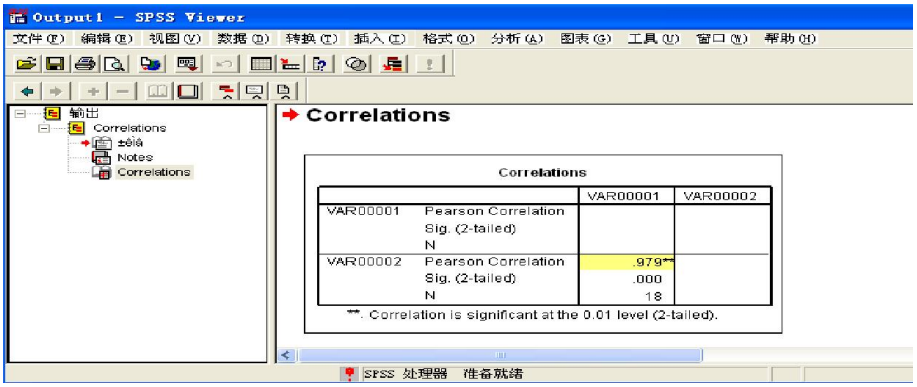


图3 相关性检验图

从图3可以看出，相关性系数为0.947，为显著相关。更清晰和详细的内容见附件。

同样，对所选取的十五个因素分别与城镇人口总数做相关性分析，结果如下表所示。

表 2 相关性分析结果

指标	相关系数	指标	相关系数
政府财政支出	0.884	教育经费	0.977
固定资产总投资	0.919	进出口总额	0.831
城市人口比重	0.993	能源生产总量	0.909
职工平均工资	0.922	第一产业产值	0.968
第二产业产值	0.930	第三产业产值	0.934
人均 GDP	0.947	社会消费品零售总额	0.983
发电量	0.950	恩格尔系数	-0.953

从表2可以看出，所选取的因素都与城镇就业人口显著相关，这是有问题的，只能说明所选取的因素与城镇就业人口趋势相同，但如果判断所选取因素与就业是否有必然联系，选取这个方法不是十分恰当，所以下面又采用了格兰杰因果关系检验的方法。

主要原因是存在伪回归的问题。我们涉及有关的时间序列都是平稳的，但是，在用一个时间序列对另一个时间序列作回归时，虽然两者之间并无任何有意义的关系，却常常会得到一个很高的R^2值，这种情况说明存在伪回归的问题。这种问题之所以产生，是因为如果所涉及的时间序列都显示出超强的趋势（持续的上升或下降），则所得到的R^2值是由于趋势的出现，而不是由于两者之间的真实关系。因此，判明经济变量之间的关系是真实的还是谬误的，是非常重要的问题，需要选择适当方法[6]。

（2）Granger 因果检验

这一问，要求从系统的角度，对可能影响到城镇就业人数与众多经济因素（比如：GDP、城镇就业人口比重、货物进出口总额、全社会固定资产投资额、政府财政支出、商品平均零售价格、通货膨胀率、第一产业生产总值、第二产业生产总值、第三产业生产总值、消费指数、发电量、钢材产量、能源生产总量、教育经费等）间的关系进行分析。

通过相关性分析只能粗略地得出与城镇就业人数密切相关的因素，考虑到这一问确立的指标将对后续问题的求解产生直接影响，运用因果关系理论更精细地分析对城镇就业人数影响最大的因素是非常必要的。

为了进行因果关系检验，先介绍一下计量经济学中经常用到的ADF单位根检验原理。

ADF 检验是在 DF 检验的基础上改进得到的，DF 检验就是先对如下回归模型进行估计：

$$\Delta y_t = \alpha + \delta y_{t-1} + u_t \qquad (1)$$

得到常规的t_δ的统计量值，然后检验假设$H_0:\delta \geq 0$;将上一步得到的t_δ值与τ统计量临界值表中的临界值进行比较，若$t_\delta > \tau$则接受假设H_0即y_t非平稳，否则，拒绝原假设，y_t为平稳序列。

由于 DF 检验中不能保证回归模型中的u_t为白噪声，所得到的估计值的不偏性难以保证。于是进行扩充形成了 ADF（Augment Dickey-Fuller）检验。

ADF 检验的回归模型有如下三种：

模型 I
$$\Delta y_t = \delta y_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta y_{t-j} + u_t \quad (2)$$

模型 II
$$\Delta y_t = \mu + \delta y_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta y_{t-j} + u_t \quad (3)$$

模型III
$$\Delta y_t = \mu + \beta t + \delta y_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta y_{t-j} + u_t \quad (4)$$

其中 p=1，2，3…模型 I 没有常数项和时间趋势，模型 II 仅含有常数项，没有时间趋势，模型III既含有常数项又含有时间趋势。

单位根的检验步骤如下：

Step 1：估计模型III。在给定的 ADF 临界检验值的显著水平下，如果δ显著地不为零，则序列不存在单位根，说明是平稳的，结束检验。否则，继续第二步。

Step 2：给定$\delta = 0$，在给定 ADF 临界值的显著水平下，如果参数β显著地不为零，则进入第三步，否则表明模型不含时间趋势，进入 step 4。

Step 3：用一般的 t 根部检验$\delta = 0$，如果参数δ显著地不为零，则序列y_t不存在单位根是平稳的，结束检验，否则，序列y_t存在单位根，是非平稳序列，结束检验。

Step 4：估计模型 II。在给定 ADF 临界值的显著水平下，如果δ显著地不为零，则序列不存在单位根，说明是平稳的，结束检验。否则，继续下一步。

Step 5：给定$\delta = 0$，在给定 ADF 临界值的显著水平下，如果参数δ显著地不为零，表明含有常数项，进入 step 3，否则进行下一步。

Step 6：估计模型 I 。在给定 ADF 临界值的显著水平下，如果参数δ显著地不为零，则序列y_t不存在单位根是平稳的，结束检验，否则，序列y_t存在单位根，是非平稳序列[7]。

例如，选定政府财政支出与城镇就业人数关系，运用 Eviews 软件进行 ADF 单位根检验[8]。

其结果如下：

Null Hypothesis: D(CP,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.411903	0.0015
Test critical values:	1% level	-4.992279	
	5% level	-3.875302	
	10% level	-3.388330	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 12

	Coefficient	Std. Error	t-Statistic	Prob.
D(CP(-1),2)	-3.466693	0.540665	-6.411903	0.0007
D(CP(-1),3)	1.579049	0.413485	3.818879	0.0088
D(CP(-2),3)	0.956952	0.295913	3.233898	0.0178
D(CP(-3),3)	0.378370	0.188838	2.003673	0.0920
C	1.548305	0.264491	5.853911	0.0011
@TREND(1990)	-0.119804	0.020773	-5.767396	0.0012

R-squared	0.928667	Mean dependent var	0.010000
Adjusted R-squared	0.869222	S.D. dependent var	0.432162
S.E. of regression	0.156284	Akaike info criterion	-0.567437
Sum squared resid	0.146547	Schwarz criterion	-0.324984
Log likelihood	9.404623	Hannan-Quinn criter.	-0.657202
F-statistic	15.62243	Durbin-Watson stat	1.596039
Prob(F-statistic)	0.002194		

图 4 ADF 单位根检验结果图

Granger 因果检验是用于检验两个变量之间因果关系的一种常用方法,于 1969 年由 J.Granger 提出,其实质是考察相互关联的两个变量之间在时间上的先导-滞后关系,即对时间序列 X_t 和 Y_t 检验以下的回归:

$$y_t = \sum_{i=1}^{q} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j y_{t-j} + u_{1t} \tag{5}$$

$$x_t = \sum_{i=1}^{s} \lambda_i x_{t-i} + \sum_{j=1}^{s} \delta_j y_{t-j} + u_{2t} \tag{6}$$

以上两个方程分别假定自身与自身的过去量和对方的过去量有关，通过对相应系数估计值的分析即可确定因果性。例如，若式***中滞后的 x 的系数估计值在统计上整体的显著不为零，同时式***中滞后的 y 的系数估计值在统计上的整体显著为零，则称 y 是引起 x 变化的原因。同理可得到其他情况的判断依据。

为了检验 x 是引起 y 的原因，Granger 因果关系检验步骤如下：

Step 1：将当前的 y 对所有滞后项 y 以及别的量做回归，但不包括滞后项 x，即受约束的回归，然后从中得到受约束残差平方和 。

Step 2：做含有滞后项 x 的回归，即在前面的回归式中加入 x 滞后项，即做无约束的回归，由此回归得到无约束的残差平方和 。

Step 3：零假设是 H_0： $\alpha_1 = \alpha_2 = \ldots = \alpha_q = 0$，即滞后项不属于此回归。

Step 4：为了检验此假设，用 F 检验，即：

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} \tag{7}$$

它遵循自由度为 q 和 $(n-k)$ 的 F 分布， n 是样本容量， q 等于滞后项 x 的个数（即有约数回归方程中待估参数的个数， k 是无约束回归中待估参数的个数。

Step 5：如果在选定的显著水平上计算的 F 值超过临界 F 值，则拒绝零假设，这样滞后的 x 项就属于此回归，这就表明 x 是 y 的原因。同样，为了检验 y 是否是 x 的原因，将 y 与 x 相互替换，重复上述步骤即可。

在应用 Granger 检验时，滞后项的个数是一个很重要的实际问题，因果方向和所含滞后项的个数有重要的关系，所以分别就不同的滞后项个数，做了因果分析，用 Eviews 软件分析得到的输出结果如下：

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 15:30
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFO does not Granger Cause LNWP	17	9.46826	0.0082
LNWP does not Granger Cause LNFO		0.02908	0.8670

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFO does not Granger Cause LNWP	16	10.3405	0.0030
LNWP does not Granger Cause LNFO		2.09627	0.1693

Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFO does not Granger Cause LNWP	15	5.12195	0.0288
LNWP does not Granger Cause LNFO		1.52235	0.2816

图 5 Granger 因果检验界面

将上图结果整理成以下表格：

表 3 Granger 因果检验结果

滞后阶数	原假设	观测值	F 值	P 值	结论
1	LnFO 不是 lnWP 的 granger 原因	17	9.46826	0.0082	拒绝
	lnWP 不是 lnFO 的 granger 原因		0.02908	0.8670	接受
2	lnFO 不是 lnWP 的 granger 原因	16	10.3405	0.0030	拒绝
	lnWP 不是 lnFO 的 granger 原因		2.09627	0.1693	接受
3	lnFO 不是 lnWP 的 granger 原因	15	5.12195	0.0288	拒绝
	lnWP 不是 lnFO 的 granger 原因		1.52235	0.2816	接受

分析 Granger 因果检验结果，在滞后阶数为 1 阶时，就可以显著地判断出政府财政支出（FO）是就业人数（WP）的 granger 原因，即政府财政支出对就业人数有着直接的快速的推动作用。而当滞后阶数为 3 时，政府财政支出仍然是就业人数的原因，即财政支出对就业人数有着长期的影响，基于此 Granger 因果检验结果，可以明显得出:政府财政支出是影响就业因素的主要指标。

用相同的方法对之前初步确定的 15 个影响就业的各项指标逐一进行分析，比较其各阶滞后条件下的 P 值。由于篇幅有限，不一一赘述。附件中含有其余 14 个指标与城镇就业人数的 Granger 因果检验的结果。

最终，本文确定了 6 个主要影响城镇就业人数的指标：人均 GDP、第三产业产值、社会零售商品总额、职工平均工资、政府财政支出以及总人口数。这样，在后续问题中，可以考虑建立城镇就业人数与这些指标之间的数学模型。

5.2 问题二的模型（Granger 因果关系理论与径向基函数神经网络的整合模型）

第二问就是要以第一问中确定的，可能影响城镇就业人数的 6 个主要指标为自变量，建立数学模型。以此来对未来就业人数的变化进行准确分析与控制，以便为未来的就业政策、现实人口政策和社会保障政策的科学制定提供有力依据。

人工神经网络（ANN）诞生于本世纪四十年代末，其结构示意图如下所示。由于它的信息分布存储、并行处理以及自学习能力的优点，已经在信息处理、模式识别、智能控制以及系统建模等领域得到了越来越广泛的应用。人工神经网络有多种（比如径向基网络，BP 网络等），本文尝试运用多种网络，以此寻找最符合解决本文就业和主要指标之间的建模问题。

径向基函数网络是一种两层前向网络，如图所示。其输出是由隐节点基函数的线性

组合计算得到。隐节点的基函数只对输入空间中的一个很小区域产生非零响应。隐层神经元将输入空间映射到一个新的空间,输出层在该空间中实现线性组合。

BP 神经网络也叫多层前馈网络，该网络的训练算法是反向传播算法，即神经元的链接权重的训练是从最后一层（输出层）开始，然后反向依次更新前一层的链接权重。任意函数都可由一个三层的网络以任意精度逼近。其两层隐藏层使用 sigmoid 单元、输出层使用非阈值的线性单元[9-11]。

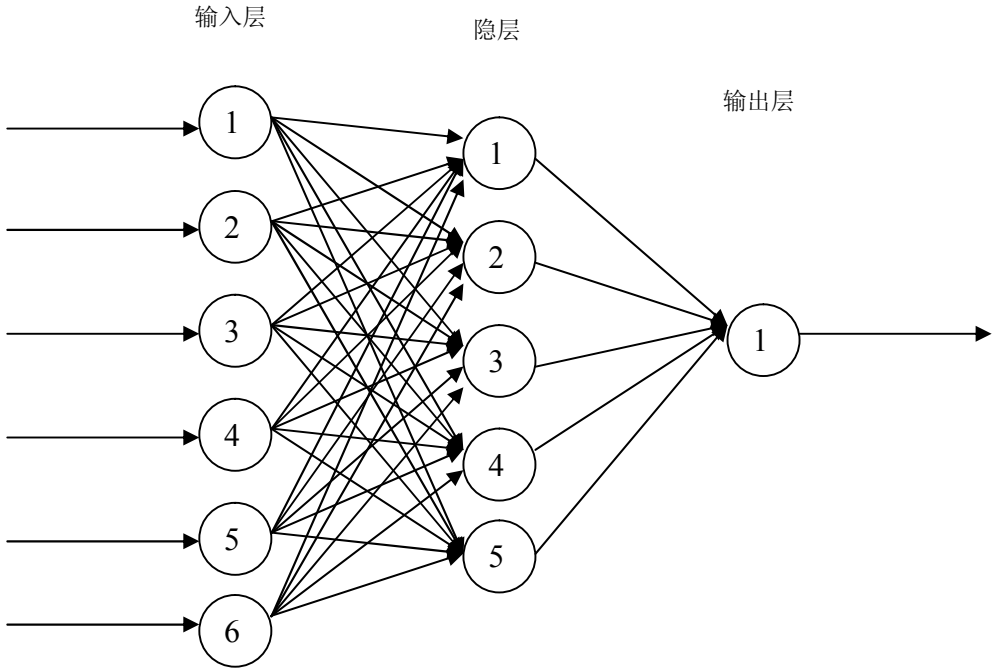


图 6 人工神经网络结构示意图[7]

借助第一问的 Granger 因果关系理论，同时利用径向基函数神经网络良好的函数逼近性，构建出因果关系理论与径向基函数神经网络的整合模型，建立起人均 GDP 等 6 个指标与城镇就业人数的定量关系。其具体实现步骤如下：

第一步：利用因果关系理论，分别对城镇就业人数及影响就业人数的各种宏观经济变量（本文全面考虑 GDP、城镇就业人口比重、货物进出口总额、全社会固定资产投资额、政府财政支出、商品平均零售价格、通货膨胀率、第一产业生产总值、第二产业生产总值、第三产业生产总值、消费指数、发电量、钢材产量、能源生产总量、教育经费等 15 项指标）进行因果关系分析，从而将通过因果检验得到的各种变量 $X_1(t), X_2(t), \cdots, X_p(t)$（本文中 $p=6$）作为径向基神经网络的输入变量，以此减少网络的输入噪音；

第二步：生成训练样本集和检验样本集。为了解决训练样本数过少和无法构建检测样本的问题，在各级指标内按随机均匀分布方式线性插值生成训练样本。考虑到数据来源不一，对欲输入的样本数据进行归一化。本文选取的归一化函数为[12]：

$$q=(x-\min)/(\max-\min) \qquad (8)$$

本文中，训练样本数据从统计年鉴中查得，对所得数据进行整理，并进行归一化处理。如下图：取 1990 年到 2005 年的 16 组作为训练样本，2006，2007 年的数据作为检验。

就业人口	人均GDP	社会消费品	职工平均工资	第三产业产	人口总数	政府财政
0	0	0	0	0	0	0
0.04120505	0.02004671	0.0189464	0.012327416	0.02144776	0.090726	0.009824
0.07968902	0.05375836	0.04575	0.035194773	0.0513581	0.172806	0.021351
0.11865889	0.10914371	0.1014038	0.075875247	0.0892347	0.254765	0.050531
0.15665695	0.19340834	0.1753297	0.14780572	0.15236398	0.335931	0.087822
0.19426628	0.27413409	0.2600987	0.207100592	0.20860395	0.413323	0.121249
0.27998056	0.33861605	0.3407149	0.250862919	0.25816809	0.490532	0.157358
0.36345967	0.38489648	0.3898465	0.26688856	0.31238283	0.565853	0.199372
0.44460641	0.41518498	0.4259424	0.329080375	0.36556735	0.634963	0.250095
0.52196307	0.44439532	0.4644943	0.382519724	0.41432012	0.697376	0.327558
0.59378037	0.50073947	0.523224	0.445697732	0.48598419	0.755648	0.41505
0.67045675	0.56231012	0.5903085	0.538091716	0.56959812	0.809475	0.512826
0.75208941	0.62487344	0.6765993	0.633752465	0.65157809	0.85977	0.614963
0.83556851	0.71705787	0.7509991	0.733481262	0.74197523	0.906899	0.699147
0.91690962	0.86159865	0.8696322	0.855769231	0.86865575	0.953236	0.823534
1	1	1	1	1	1	1
1.09514091	1.17019905	1.1568266	1.162536982	1.1671275	1.042136	1.210475
1.19620991	1.39334354	1.3742308	1.404832347	1.39412026	1.083602	1.513866

图 7 训练样本数据

第三步：确定城镇就业人数 $Y(t)$ 为网络的输出变量；

第四步：确定隐层节点数目。隐层节点数由经验和试验决定，参考经验公式 $2^m > N$（m：隐层节点数，N 为样本数）。

第五步：尝试运用不同种类的神经网络，寻找最符合解决本文就业和主要指标之间的建模问题的网络。

① 用 Matlab 径向基神经网络软件包，用已有的 17 组数据中的十五组做训练样本，另两组作检验样本，得到的检验结果十分不理想，排除此法。

② 用 Matlab 的 BP 神经网络软件包，此时隐层神经元数在 4 左右穷举，可知 5 个神经元时收敛最快，即为典型的 6x5x1 拓扑结构。训练误差设置为 0.0001，最大迭代步数是 500 次。

③ 考虑到 BP 神经网络固有的缺陷（比如：训练次数多，学习效率低，收敛速度慢，易形成局部最优而得不到全局最优等），采用改进的 BP 神经网络。本文采用网络训练方法，采用改进 BP 算法中的高斯-牛顿法。以下是这两种网络具体求解过程的比较。

训练过程的比较：

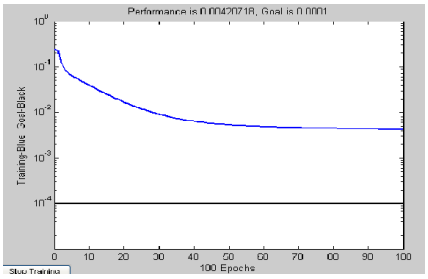


图 8 BP 神经网络训练过程

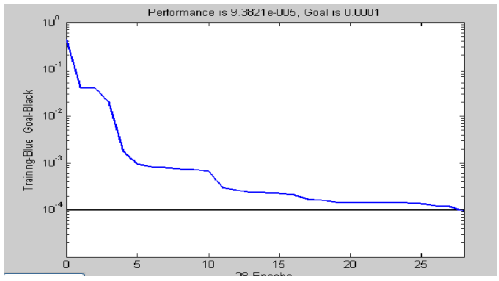


图 9 改进 BP 神经网络训练过程

同时，还得到了各自的数据拟合情况比较：

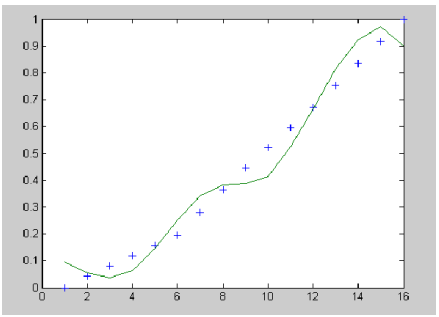


图 10 BP 神经网络拟合

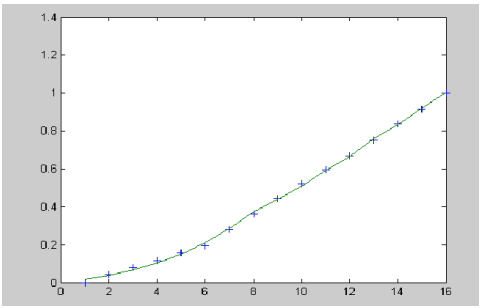


图 11 改进神经网络拟合

由以上 4 幅图比较 BP 神经网络和改进的 BP 神经网络，可见，改进的 BP 神经网络的逼近误差是非常小的，收敛效果相当好。

至此，很明显得出：径向基网络，BP 神经网络以及改进的 BP 神经网络中，改进的 BP 神经网络是最适合的方法。

另外，本文还用改进的 BP 神经网络解得输入层的最终权值矩阵为：

$$\begin{pmatrix} 1.2004 & -1.4294 & -1.7494 & 1.9491 & 1.7499 & 0.0244 \\ -0.5487 & -2.2736 & -1.5645 & -1.5981 & -0.8094 & -1.5098 \\ -0.5266 & 2.6166 & -0.8338 & -1.0425 & 2.1096 & 0.2231 \\ -0.2303 & 0.1850 & -1.2539 & -2.5683 & -1.2111 & 1.9194 \\ 0.4917 & -1.3044 & -2.0346 & -1.5077 & -0.4558 & -2.2006 \end{pmatrix}$$

进而对每一列求代数和，并取绝对值可得 6 个主要指标的权值表：

人均国内生产总值	社会消费品零售总额	职工平均工资	第三产业产值	人口总数	政府财政支出
0.3865	2.2058	7.4362	4.7675	1.3832	1.5435

5.3 问题三的模型

由第 2 节对第三问的问题分析可知，此小节是本文的重点所在。本节要完成的主要任务如下：

第一，建立主要经济社会指标与就业人数的可加性分解模型。

第二，建立第一、第二、第三产业与就业人数的 VEC 模型。

第三，利用主成分分析法和层次聚类法，对全国进行区域划分。

第四，利用最小二乘法拟合大学生就业率，并对未来两年进行预测。

（1） 主要经济社会指标与就业的关系（可加性分解模型）

本小节重点选定经济社会指标中的城镇人均 GDP、经济增长率、产业结构变化、进出口总额、中间需求（与消费品的数量有关）、国内需求等指标，从投入产出经济理论出发，建立城镇就业人数与上述选定指标之间的定量数学模型。

令 $I_t = L_t / X_t$ 是 t 时期的工作产出率，则有：

$$L_t = I_t X_t \tag{9}$$

其中，L_t 为社会所能容纳的总就业人数，它在 t 时期的工作产出为 X_t 。那么，由此可以得出社会在 t 到 $t+1$ 时期的变化为：

$$\begin{aligned} \Delta L_t &= I_{t+1}X_{t+1} - I_t X_t \\ &= X_{t+1}\Delta I_t + I_t \Delta X_t \end{aligned} \tag{10}$$

上式就是用就业系数和产出加权的就业人数变化总量，其权重分别为期末产量和其初就业系数。上式还可以改写成：

$$\begin{aligned} \Delta L_t &= I_{t+1}X_{t+1} - I_{t+1}X_t + I_{t+1}X_t - I_t X_t \\ &= X_t \Delta I_t + I_{t+1}\Delta X_t \end{aligned} \tag{11}$$

其权重分别为期初产量和期末就业系数。

为了消除权重偏差，按照马歇尔-艾奇渥斯指数加权平均，得到第一分解模型为：

$$\Delta L_t = 0.5(X_t + X_{t+1})\Delta I_t + 0.5(I_t + I_{t+1})\Delta X_t \tag{12}$$

此式表明ΔL_t可分解为就业系数和产出变化而引起的变化之和。具体可解释为：

① 给定生产水平，就业系数的变化带来就业人数吸收和工作效率的变化，此为技术效应；

② 给定工作效率,则当产出变化时，就业也将会随之发生变化，此为产出效应。

式（12）是源于独立模型的分解式，与复杂的实际情况有相当大的差距。下面利用投入产出模型推导出分解方案。

在良性静态开放投入产出系统中，存在平衡关系式：

$$X = AX + Y + E - M \tag{13}$$

上式中，X,Y,E,M代表总产出、最终需求、出口和竞争性进口的n维列向量；A为直接消耗系数矩阵。记

$$D = Y + E - M \tag{14}$$

则有

$$X = (I - A)^{-1} D \tag{15}$$

写成分量形式，即为：

$$X_i = \sum_{j=1}^{n} r_{ij} d_j \qquad (i = 1, 2, \cdots, n) \tag{16}$$

上式中，r_{ij}为 Leontief 矩阵$(I - A)^{-1}$中第i行第j列的元素。

将上式（16）代入式（9），可以推出：

$$L_t = I_t \sum_{j=1}^{n} r_{ij} d_j \qquad (i = 1, 2, \cdots, n) \tag{17}$$

将上式（17）关于时间求导数，则有：

$$\dot{L}_t = \dot{I}_t \sum_{j=1}^{n} r_{ij} d_j + I_t \sum_{j=1}^{n} \dot{r}_{ij} d_j + I_t \sum_{j=1}^{n} r_{ij} \dot{d}_j \tag{18}$$

同理，利用马歇尔-艾奇渥斯指数加权平均，上式（）可以表示为第二分解模型：

$$\begin{aligned} \Delta L_t &= 0.5 \Delta I_t \sum (r_{ij,t+1} d_{j,t+1} + r_{ij,t} d_{j,t}) \\ &+ 0.5 (I_{i,t+1} + I_{i,t}) \sum r_{ij,t} d_{j,t} + 0.5 (I_{i,t+1} + I_{i,t}) \sum r_{ij,t+1} d_{j,t} \end{aligned} \tag{19}$$

该式表明，就业随就业系数、结构性相关和最终需求的变化而变化，后两个因素在式（）中式隐含的。

上式的三因素分解还可以进一步扩展。事实上，最终需求和就业系数部分还可以进一步细分，从而产生更加细化的模型。

因为

$$D = Y + E - M \tag{20}$$

那么，可以得出：

$$\Delta d_{j,t} = \Delta Y_{j,t} + \Delta e_{j,t} - \Delta m_{j,t} \tag{21}$$

如果定义

$$z_j = x_j + m_j, u_j = m_j / z_j \tag{22}$$

由此有

$$m_j = u_j z_j \tag{23}$$

并且，更进一步，可以推出：

$$\Delta m_{j,t} = 0.5 \Delta u_{j,t} (z_{j,t+1} + z_{j,t}) + 0.5 (u_{j,t+1} + u_{j,t}) \Delta z_{j,t} \tag{24}$$

代入到投入产出模型中，得到：

$$z_j = x_j + m_j = w_j + y_j + e_j \qquad w_j = \sum_{i=1}^{n} a_{j,i} x_i \tag{25}$$

则有

$$\Delta z_j = \Delta w_j + \Delta y_j + \Delta e_j \tag{26}$$

将式（25）（26）代入式（21），得：

$$\begin{aligned}\Delta d_{j,t} &= 0.5\Delta u_{j,t}(z_{j,t+1} + z_{j,t}) - 0.5(u_{j,t+1} + u_{j,t})\Delta w_{j,t} \\ &+ (1 - 0.5(u_{j,t+1} + u_{j,t}))(\Delta y_{j,t} + \Delta e_{j,t})\end{aligned} \tag{27}$$

同样，定义 $s_i = L_i / K_i$ 和 $k_i = K_i / X_i$，所以有

$$I_i = s_i k_i \tag{28}$$

从而：

$$\Delta I_{i,t} = 0.5\Delta s_{i,t}(k_{i,t+1} + k_{i,t}) + 0.5(s_{i,t+1} + s_{i,t})\Delta k_{i,t} \tag{29}$$

这表明就业系数随劳动资本率 s_i 和资本系数 k_i 的变化而变化。

将式（28）（29）代入到式（19），就会推导出一个有七部分的第三分解模型：

$$\begin{aligned}\Delta L_t &= 0.25\Delta s_{i,t}(k_{i,t+1} + k_{i,t})\sum(r_{ij,t+1}d_{j,t+1} + r_{ij,t}d_{j,t}) \\ &+0.25(s_{i,t+1} + s_{i,t})\Delta k_{i,t}\sum(r_{ij,t+1}d_{j,t+1} + r_{ij,t}d_{j,t}) \\ &+0.5(I_{i,t+1} + I_{i,t})\sum\Delta r_{ij,t}d_{j,t} \\ &-0.25(I_{i,t+1} + I_{i,t})\sum r_{ij,t+1}\Delta u_{j,t}(z_{j,t+1} + z_{j,t}) \\ &-0.25(I_{i,t+1} + I_{i,t})\sum r_{ij,t+1}(u_{j,t+1} + u_{j,t})\Delta w_{j,t} \\ &+0.25(I_{i,t+1} + I_{i,t})\sum r_{ij,t+1}[2 - (u_{j,t+1} + u_{j,t})]\Delta y_{j,t} \\ &+0.25(I_{i,t+1} + I_{i,t})\sum r_{ij,t+1}[2 - (u_{j,t+1} + u_{j,t})]\Delta e_{j,t}\end{aligned} \tag{30}$$

至此，便得到了一个包含七部分的可加性分解模型。

第一部分度量了人均 GDP 变化的影响；

第二部分度量了资本产出率变化出率变化的影响；

第三部分显示了产业间连锁变化的影响；

第四部分为是进口系数变化的影响，展示了进口替代的影响（当进口替代发生，即 $u_j < 0$ 时，国内生产和由此而来的就业便会增长，进口替代的直接与间接影响均已考虑了）；

第五部分展示了对产出的中间需求变化的影响，对任何产出的中间需求降低，将意味着具有重要意义的消费品增长，并随之将可能产生更多的就业机会；

第六部分为国内最终需求变化的影响；

第七部分为出口变化的影响。因为 $u_j < 1$，所以对每个 j 有 $2 - (u_{j,t+1} + u_{j,t}) > 0$，第六、七部分的符号将分别由 $\Delta y_{j,t}$ 和 $\Delta e_{j,t}$ 的符号来决定[13]。

（2）产业（行业）与就业关系（VEC 模型）

根据第二节的问题分析，把行业的影响理解为产业结构对城镇就业人数的影响。产业结构，亦称国民经济的部门结构。国民经济各产业部门之间以及各产业部门内部的构成。它反映了各个生产要素在国民经济各产业部门之间的比例构成以及它们之间的相互依存的关系。

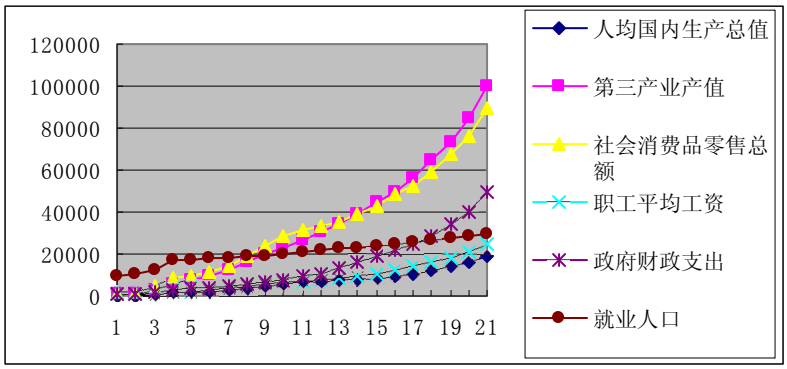


图 12 各年份第一二三产业与 GDP 比值趋势；第一二三产业就业人数比重趋势

发达国家产业结构的演进历程表明，伴随着经济水平的不断提高，劳动力在各产业之间将发生规律性转移。首先是劳动力从第一产业转向第二产业，随着经济水平的进一步提高，劳动力将更多地向第三产业转移。目前，发达国家第三产业的就业比重已超过60%，第三产业已成为发达国家就业的主要行业。第三产业就业比重的提高不仅是吸纳剩余劳动力的需要，更是产业结构高级化的要求[14]。改革开放以来，中国经济发展与产业结构、就业结构发生了怎样的变化，产业结构与就业结构是否协调、二者之间呈现何种定量关系？为此，建立如下 VEC 数学模型。

首先考虑经济学中的 VAR 模型（向量自回归模型）。在经济活动中，广泛存在着时间滞后效应，即动态性。在经济活动中，广泛存在着时间滞后效应，即动态性。某些经济变量不仅受到同期各种因素的影响，而且也受到过去某些时期的各种因素甚至自身的过去值的影响。通常把这种过去时期的具有滞后作用的变量叫做滞后变量。造成滞后变量的原因很多，主要有心理原因，技术原因，制度原因等。

一般的 VAR 模型为：

$$Y_t = \alpha + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t \tag{31}$$

上式中，Y_t 是由内生变量组成的向量，即 Y_t=（就业结构，产业结构），P 为滞后项数，$A_1, A_2, \cdots A_p$ 为参数矩阵，ε_t 为扰动向量，可视为白噪声过程，且满足：

$$E(\varepsilon_{i,t}, \varepsilon_{i,t}) = 0 (i, j = 1, 2, \cdots, n, i \neq j) \tag{32}$$

下面，在 VAR 模型的基础上，建立 VEC 模型（向量误差修正模型）。在建模之前，先说明一下协整关系的含义，协整关系作为经济术语之一，指变量之间存在的长期均衡关系。在 VAR 模型中，每个方程都是一个自回归分布滞后模型，而 VEC 模型可以看作是含有协整约束的 VAR 模型[15]。其表述如下：

$$\Delta Y_t = \alpha K_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \varepsilon_t, t = 1, 2, \cdots T \tag{33}$$

其中，

$$K_{t-1} = \beta Y_{t-1}$$

上式（33）中，Y_t 包含 k 个存在协整关系的 I(1) 变量，但不包含外生变量，其中每个方程的误差项都具有平稳性，误差修正项反映变量之间的长期均衡关系。

本文首先对六个变量进行单整检验。x_i，第 i 产业的产业占 GDP 的比重，y_i：第 i 产业的就业比重。具体结果如下：

表 4 产业单整检验结果

待检验指标	差分次数	F 值	P 值	结论
第一产业产值比重	1	-4.109831	0.0162	一阶单整
第二产业产值比重	1	-5.383132	0.0049	一阶单整
第三产业产值比重	1	-4.392166	0.0100	一阶单整
第一产业就业比重	1	-5.266759	0.0015	一阶单整
第二产业就业比重	1	-4.064857	0.0201	一阶单整
第三产业就业比重	1	-3.929904	0.0265	一阶单整

由上述结果知，第一产业 $x_1 \sim$ I（1），$y_1 \sim$ I（1），第二产业 $x_2 \sim$ I（1），$y_2 \sim$ I（1），第三产业 $x_2 \sim$ I（1），$y_2 \sim$ I（1），即说明各产业就业比重与产业比重都是一阶单整的，在此基础上再做协整检验，采用 JJ 检验可得（篇幅关系，仅列出一个检验结果图，其他结果图详见附件）：

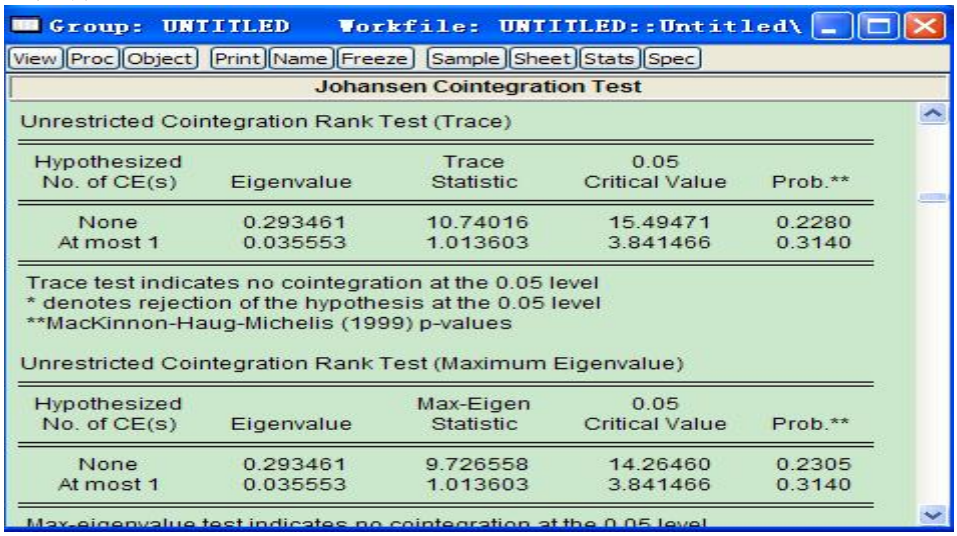


图 13 JJ 检验结果图

经过检验，可以看出：第一产业，该检验选择的滞后阶数为 4 时，y_1 与 x_1 存在协整关系；对于第二产业，该检验选择的滞后阶数为 4 时，y_2 与 x_2 存在协整关系；对于第三产业该检验选择的滞后阶数为 5 时，y_3 与 x_3 存在协整关系。

据此，我们可以分别建立三次产业就业结构与产业结构的 VEC 模型。

第一产业 VEC 模型：

$$\begin{bmatrix}\Delta y_t\\ \Delta x_t\end{bmatrix}=\begin{bmatrix}-1.329\\ 0.637\end{bmatrix}+\begin{bmatrix}-1.323\\ 0.637\end{bmatrix}ecm_{t-1}+\begin{bmatrix}0.533 & 0.045\\ -0.408 & 0.529\end{bmatrix}\begin{bmatrix}\Delta y_{t-1}\\ \Delta x_{t-1}\end{bmatrix}$$

$$+\begin{bmatrix}-0.152 & 0.246\\ 1.019 & 0.083\end{bmatrix}\begin{bmatrix}\Delta y_{t-2}\\ \Delta x_{t-2}\end{bmatrix}+\begin{bmatrix}0.068 & -0.184\\ 0.167 & 0.123\end{bmatrix}\begin{bmatrix}\Delta y_{t-3}\\ \Delta x_{t-3}\end{bmatrix}+\begin{bmatrix}-0.572 & -0.222\\ 0.191 & 0.116\end{bmatrix}\begin{bmatrix}\Delta y_{t-4}\\ \Delta x_{t-4}\end{bmatrix}$$

$$ecm_{t-1}=y_{t-1}+1.444x_{t-1}-103.764$$

t=23.3763 R^2=0.062

第二产业 VEC 模型：

$$\begin{bmatrix}\Delta y_t\\ \Delta x_t\end{bmatrix}=\begin{bmatrix}0.679\\ 0.164\end{bmatrix}+\begin{bmatrix}0.004\\ 0.074\end{bmatrix}ecm_{t-1}+\begin{bmatrix}0.487 & 0.157\\ -0.018 & 0.649\end{bmatrix}\begin{bmatrix}\Delta y_{t-1}\\ \Delta x_{t-1}\end{bmatrix}$$

$$+\begin{bmatrix}-0.138 & -0.169\\ -0.075 & 0.395\end{bmatrix}\begin{bmatrix}\Delta y_{t-2}\\ \Delta x_{t-2}\end{bmatrix}+\begin{bmatrix}-0.054 & 0.085\\ 0.033 & -0.129\end{bmatrix}\begin{bmatrix}\Delta y_{t-3}\\ \Delta x_{t-3}\end{bmatrix}+\begin{bmatrix}-0.134 & -0.169\\ -0.224 & 0.035\end{bmatrix}\begin{bmatrix}\Delta y_{t-4}\\ \Delta x_{t-4}\end{bmatrix}$$

$$ecm_{t-1}=y_{t-1}-3.935x_{t-1}+63.945$$

t=-3.701　　R^2=1.063

第三产业 VEC 模型：

$$\begin{bmatrix}\Delta y_t\\ \Delta x_t\end{bmatrix}=\begin{bmatrix}0.057\\ 0.117\end{bmatrix}+\begin{bmatrix}-1.557\\ -1.653\end{bmatrix}ecm_{t-1}+\begin{bmatrix}0.956 & 1.973\\ -0.249 & 0.085\end{bmatrix}\begin{bmatrix}\Delta y_{t-1}\\ \Delta x_{t-1}\end{bmatrix}+\begin{bmatrix}0.777 & -0.198\\ -0.035 & 0.094\end{bmatrix}\begin{bmatrix}\Delta y_{t-2}\\ \Delta x_{t-2}\end{bmatrix}$$

$$+\begin{bmatrix}0.437 & -0.264\\ -0.233 & 0.011\end{bmatrix}\begin{bmatrix}\Delta y_{t-3}\\ \Delta x_{t-3}\end{bmatrix}+\begin{bmatrix}0.807 & 1.519\\ -0.103 & -0.305\end{bmatrix}\begin{bmatrix}\Delta y_{t-4}\\ \Delta x_{t-4}\end{bmatrix}+\begin{bmatrix}0.249 & 0.737\\ -0.162 & -0.313\end{bmatrix}\begin{bmatrix}\Delta y_{t-5}\\ \Delta x_{t-5}\end{bmatrix}$$

$$ecm_{t-1}=y_{t-1}-0.189x_{t-1}-1.315$$

t=-31.273　　R^2=0.001

（3）地区与就业的关系（区域划分）

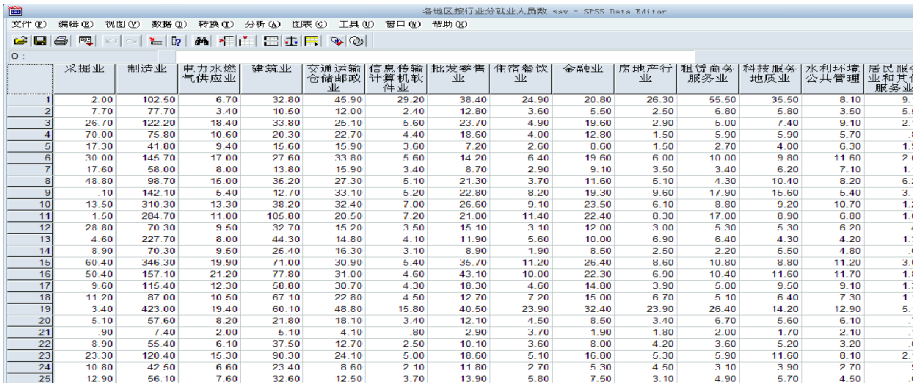
就业问题是关系国计民生的重大问题，。目前，中国面临就业的人口数量和增长速度均居世界首位，预计 2020 年劳动年龄总人口将达到 9.4 亿人。庞大的就业人口基数对中国经济社会和谐发展构成了巨大的压力。

可见，本文将要探讨的中国就业的空间规律,由此形成的就业区域划分，对政府流动人口的引导和就业政策制订具有重要的现实意义。

① 数据来源

根据 2000 年第五次《人口普查分县资料》，可以对中国内地 2869 个县（市）、区（不涉及港澳台地区）的数据进行分析和研究[16]。但鉴于 2000 年距今由于时间上的差异以及经济的变革，很多需要的数据都产生了很大的变化，并且数据量也非常大，不易短时间内完成，所以参考中国国家统计局《中国统计年鉴 2008》，按全国的行政区划进行数据的分析。由于要揭示中国的就业空间规律,需要通过确立有代表意义的指标体系和指标项,并运用定量分析方法进行探索。根据中国国家统计局《中国统计年鉴 2008》对行业的划分，共分为 19 个门类：农林牧渔业,采矿业,制造业,电力，燃气及水的生产和供应业,建筑业,交通运输、仓储和邮政业,信息传输、计算机服务和软件业，批发和零售业,住宿和餐饮业，金融业,房地产业,租赁和商务服务业,科学研究、技术服务和地质勘查业，水利、环境和公共设施管理业，居民服务和其他服务业，卫生、社会保障和社会福利业,教育、文化体育和娱乐业,公共管理和社会组织，本文选取了全国行政区划的上述 19 个行业指标。这些行业全面反映了各个统计单位第一、第二、第三产业的就业结构。

本文中，选取的原始矩阵为行 19 列的矩阵，利用 Eviews 软件进行求解，输入原始矩阵的部分界面如图 14 所示。



	采掘业	制造业	电力水燃气供应业	建筑业	交通运输仓储邮政业	信息传输计算机软件业	批发零售业	住宿餐饮业	金融业	房地产行业	租赁商务服务业	科技服务地质业	水利环境公共管理	居民服务和其他服务业
1	2.00	102.50	6.70	32.80	45.90	29.20	38.40	24.90	20.80	26.30	55.50	35.50	8.10	9.
2	7.70	77.70	3.40	10.60	12.00	2.40	12.80	3.60	5.50	2.50	6.80	5.80	3.50	5.
3	26.70	122.20	18.40	33.80	26.10	5.60	23.70	4.90	19.60	2.90	5.00	7.40	9.10	2.
4	70.00	75.80	10.60	20.30	22.70	4.40	18.60	4.00	12.80	1.50	5.90	5.90	5.70	.
5	17.30	41.60	9.40	15.60	15.90	3.00	7.20	2.00	0.00	1.50	2.70	4.00	6.30	1.
6	30.00	145.70	17.00	27.60	33.80	5.60	14.20	6.40	19.60	6.00	10.00	9.80	11.60	2.
7	17.60	58.00	8.00	13.80	15.90	3.40	8.70	2.90	9.10	3.50	3.40	6.20	7.10	1.
8	18.80	98.70	16.00	36.20	27.30	5.10	21.30	3.70	11.60	5.10	4.30	10.40	8.20	6.
9	1.10	142.10	5.40	12.70	33.10	5.20	22.80	8.20	19.30	9.60	17.90	15.60	5.40	3.
10	13.50	316.30	13.30	38.20	32.40	7.00	26.60	9.10	23.50	6.10	8.80	9.20	10.70	1.
11	1.50	284.70	11.00	105.00	20.50	7.20	21.00	11.40	22.40	0.30	17.00	8.90	6.00	1.
12	28.80	70.30	9.50	32.70	15.20	3.50	15.10	3.10	12.00	3.00	5.30	5.30	6.20	.
13	4.60	227.70	8.00	44.30	14.80	4.10	11.90	5.60	10.00	6.90	6.40	4.30	4.20	1.
14	8.90	70.30	9.50	26.10	16.30	3.10	8.90	1.90	8.50	2.50	2.20	5.50	4.80	.
15	60.40	346.30	19.90	71.00	30.90	5.40	35.70	11.20	26.40	8.60	10.80	8.80	11.20	3.
16	50.40	157.10	21.20	77.80	31.00	4.60	43.10	10.00	22.30	6.00	10.40	11.60	11.70	1.
17	9.60	115.40	12.30	50.00	30.70	4.30	18.30	4.60	14.00	3.90	5.00	9.50	9.10	1.
18	11.20	87.00	10.50	67.10	22.80	4.50	12.70	7.20	15.00	6.70	5.10	6.40	7.30	1
19	3.40	423.00	19.40	60.10	48.80	15.80	40.50	23.90	32.40	23.90	26.40	14.20	12.90	5.
20	5.10	57.60	8.20	21.80	18.10	3.40	12.10	4.60	8.50	3.40	6.70	5.60	6.10	.
21	.90	7.40	2.00	5.10	4.10	.80	2.90	3.70	1.90	1.80	2.00	1.70	2.10	.
22	8.90	55.40	6.10	37.50	12.70	2.50	10.10	3.60	8.00	4.20	3.60	5.20	3.20	.
23	23.30	120.40	15.30	90.30	24.10	5.00	18.60	5.10	16.00	5.30	5.90	11.00	8.10	2.
24	10.80	42.50	6.60	23.40	8.60	2.10	11.80	2.70	5.30	4.50	3.10	3.90	2.70	.
25	12.90	56.10	7.60	32.60	12.50	3.70	13.90	5.80	7.50	3.10	4.90	5.70	4.50	.

图 14 SPSS 的输入矩阵界面截图

② 数据处理（主成分分析和聚类方法定量分析）

本文以每个全国行政区划为地理统计单元。取19个行业的就业人口空间密度和就业行业百分比为指标（在后面的主成分分析中，为了解决指标间数量级的差异悬殊以及为了研究的直观与方便，取该地区就业人口数为指标并进行标准化），采用主成分分析方法和层次聚类方法, 对各地理单元的就业空间结构及类型进行归类。

主成分分析（即 PCA）：是最古老的多元统计分析技术之一，可将数据从高维数据空间变换到低维特征空间，因而可以用于数据特征提取，其实质是将研究对象的多个相关变量转化为少数几个不相关变量的一种多元统计方法[17]。

它基于 Karhunen-Loeve 分解，目的是在数据空间中找一组向量尽可能的解释数据方差，通过一个特殊的向量矩阵，将数据从原来的高维空间映射到一个低维空间，降维后保存了数据的主要信息，从而简化数据处理过程。

选择所确立的宏观经济指标作为样品的原始数据组成矩阵，设有 N 个地区，对它们各观测 P 个指标变量，具体来说，对每个地区选取就业人口总数，其原始矩阵为：

$$X=\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & & \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \tag{34}$$

其中 x_{ij} 表示第 i 个地区的第 j 个指标的值（$i=1,2,\cdots,n; j=1,2,\cdots,p$）。

主要步骤为：

Step 1：对原始数据进行标准化处理，形成标准化矩阵：

$$x'_{ij}=\frac{x_{ij}-\overline{x}_j}{s_j} \qquad (i=1,2,\cdots,n) \tag{35}$$

其中，均值 $\overline{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij} \qquad (j=1,2,\cdots,p)$

方差

$$s_j=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2} \qquad (j=j=1,2,\cdots,p) \tag{36}$$

假定经过变换后为 X，则 X 的元素 x_{ij} 的均值 0，方差为 1，各单项指标具有相同度量尺度和一致的变化范围。

Step 2：计算相关矩阵及其特征根，选取主成分；

相关矩阵 R 是一个对称矩阵 $R=(r_{ij})$，

其中

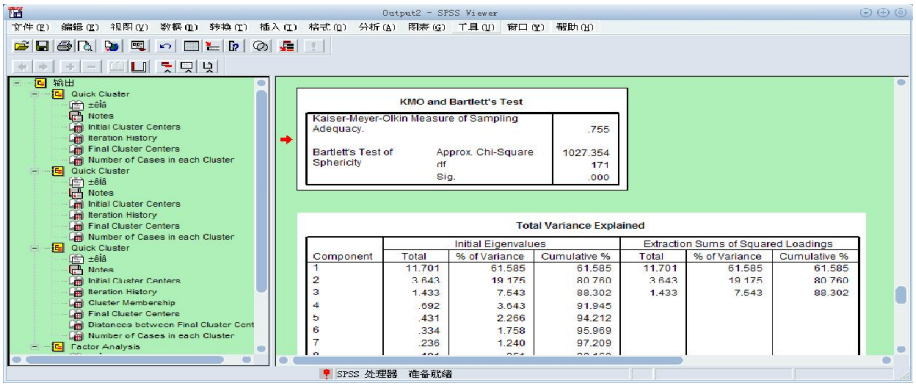
$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{kj}-\bar{x}_i)(x_{kj}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{kj}-\bar{x}_i)^2}\sqrt{\sum_{k=1}^{n}(x_{kj}-\bar{x}_j)^2}}(i=1,2,\cdots,n;j=1,2,\cdots,p) \tag{37}$$

Step 3：在 step 2 的基础上利用雅可比法求 R 的全部特征根 λ_i，并由大到小排列 $\lambda_1>\lambda_2\cdots>\lambda_p>0$，再求相应的正则化单位特征向量 $\xi_i=(\xi_{1i},\xi_{2i},\cdots\xi_{pi})$，则第 i 个主成分表示为各个指标 X_k 的线性组合 $Z_i=\sum_{k=1}^{p}l_{ki}X_k$；

Step 4：确定主成分数目。在确定主成分数目之前，需要先给出一个控制值 α，令 $\sum_{i=1}^{q}\lambda_i/\sum_{i=1}^{p}\lambda_i\geq 1-\alpha$，则对应满足条件的 q 的最小值即为保留的主成分的个数 m，本文取 $\alpha=0.05$；

Step 5： 计算综合得分。首先，计算得到第 i 个样本中第 k 个主成分的得分为 $F_{ik}=\sum_{j=1}^{p}l_{jk}X_j$，再以 m 个主成分的方差贡献率为权重求得第 i 个样本的综合得分 $f_i=\sum_{j=1}^{p}F_{ik}\times\lambda_k$，并根据每个样本的综合得分，进行排序。

基于如上步骤进行主成分分析后，得到如下的结果：



KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.755
Bartlett's Test of Sphericity	Approx. Chi-Square	1027.354
	df	171
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.701	61.585	61.585	11.701	61.585	61.585
2	3.643	19.175	80.760	3.643	19.175	80.760
3	1.433	7.543	88.302	1.433	7.543	88.302
4	.692	3.643	91.945			
5	.431	2.266	94.212			
6	.334	1.758	95.969			
7	.236	1.240	97.209			

图 15 方差解释及因子提取结果图

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component		
	1	2	3
农林牧渔业	-.005	.172	.835
采掘业	.293	.584	.503
制造业	.786	.217	-.286
电力水燃气供应业	.775	.676	.139
建筑业	.672	.306	-.296
交通运输仓储邮政业	.947	-.069	.133
信息传输计算机软件业	.777	-.584	.046
批发零售业	.948	.041	.028
住宿餐饮业	.884	-.382	-.143
金融业	.940	.165	.117
房地产行业	.825	-.496	-.089
租赁商务服务业	.734	-.653	-.021
科技服务地质业	.770	-.530	.160
水利环境公共管理	.884	.335	.148

图 16 因子载荷矩阵

具体内容见附件。

其中，选择特征值在1.0以上，KMO检验值为0.755，各变量间的相关程度无太大差别，主成分有意义。运算后，前三个主成分的累计方差可达88.302%，第1、2、3 主成分贡献率分别为61.585%，19.175%，7.543%，并得出3 个主成分及其与19个就业指标的

关系(如图16或附件)。
第1主成分中,得分较高的指标有交通运输、仓储及邮电通信业，批发零售业，餐饮业， 金融业，房地产业，水利环境公共管理业，卫生社保福利业，文体娱乐业等，可概括为三产就业主成分；第2主成分中，得分较高的指标有电力燃气及水的生产和供应业、建筑业、地质业及其它行业得分高，属于二产就业主成分；第3主成分中，农林牧渔业的得分最高，属于一产就业主成分。

聚类分析：运用回归法计算出地区的3 个主成分得分（本文的情况也可以在已有的统计找到，经验证，计算出的结果与所查数据基本上一致）。再通过层次聚类法将其聚为3 类，分析各地区与类中心的空间位置，可以总结为：第三产业主导就业地区，主要是直辖市及主要省会、开放城市等（如北京、上海比重很大）；第二产业主导就业地区，布局相对集中, 主要在我国东部沿海地区和东北沿边地区，大城市周围的中小城市及矿产采掘业就业区；第一产业主导就业地区，主要分布在第二、三产业主导就业区的外围区，在我国就业空间体系上占有主导地位。

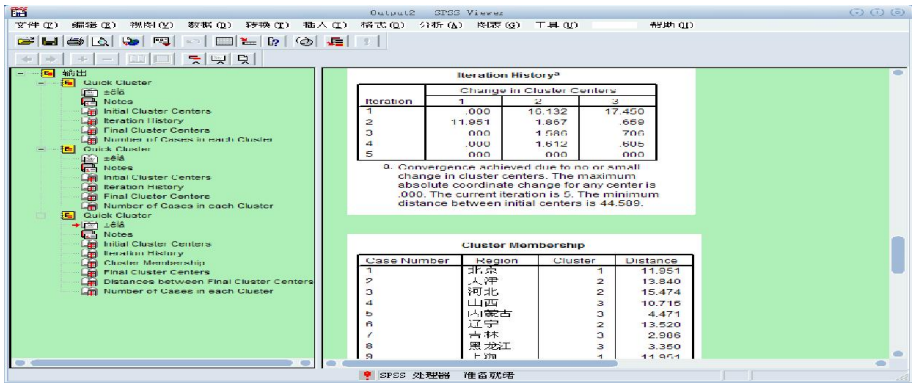


图17 地区所属类分布结果

其中图 16 界面中的 VAR00001，VAR00002 和 VAR00003 分别表示第三、第二和第一产业。 利用这类模型（建立地区与就业指标的关系矩阵，通过主成分分析法和聚类分析的方法求解，得到地区在就业模式上的划分），同样适用于求解其他范围的区域如黑龙江省，或东北地区等。

明确了地区为哪一个产业为主导，还可以通过上述产业与就业关系的模型进一步求解。

（4）人群与就业的关系（仅建立大学生就业率模型）

对就业人群，可能由于受教育程度、专业技能的掌握程度、个人工作能力、环境适应能力等众多因素的不同，都会对就业人数造成一定影响。本文选择从整体上对就业人群进行把握，抛开其他次要因素，仅考虑受教育程度的差异而导致人群就业率的不同[]。

下图为我国人民受教育程度统计图：

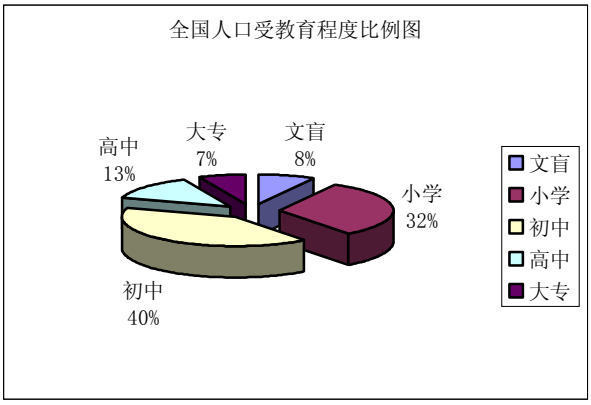


图 18 全国人口受教育程度比例图

随着高等教育的迅猛发展，“大学生就业率”是我国高校所面临的重要问题，而它也是与同学们密切相关的一个实际问题。大学生就业率能够更科学、更准确地反映出高校的实际就业状况和办学水平。并且，也反映了高校教育和培养工作与社会需求的适应程度，关系到高校的生存与发展，同时对考生、家长乃至整个社会都具有很大的影响。作为评价一所高校办学水平的重要指标，它在一定程度上对高校调整专业结构、提高办学质量、重视就业工作起到了积极的促进作用。

毕业生就业指导工作是一项繁琐而又需细致的工作，单纯的手工操作和经验管理不仅工作量大，耗费人力多，而且效率低，还比较容易出差错。对于学生处和各个学院来说，这样做的结果往往是费时，费力，不利于动态掌握学生的就业信息，不利于及时向有关部门反映情况；对上级部门来说，也不能在第一时间把握最新的信息，从而影响到下一步的工作部署。

本文以年份作为函数的自变量用 x 表示，由于就业是一个动态过程，随着月份的不同，就业率将不断上升。本文选定黑龙江省每年 9 月和年底教育部门公布的大学生就业率[19,20]作为变量用 $f(x)$ 表示。

下图体现了 2004 年至 2008 年 5 年间黑龙江省大学生就业率变化趋势：

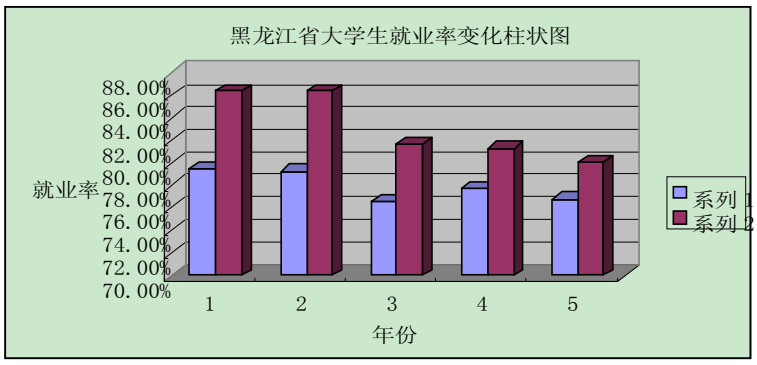


图 19 黑龙江大学生就业率变化柱状图

建立模型如下：

$$\sum_{j=0}^{n}\rho(x_j)[f(x_j)-s^*(x_j)]^2=\min_{s\in\Phi}\sum_{j=0}^{m}\rho(x_j)[f(x_j)-s(x_j)]^2 \tag{38}$$

其中，$\rho(x_j)$ 为权值，并且 $\sum_{j=0}^{n}\rho(x_j)\equiv 1$，$s^*(x_j)$ 为通过就业数据拟合的曲线，Φ 为线性无关的函数类集合。由于 $s\in\Phi$，所以 $s(x)$ 可以表示为：

$$s(x)=\sum_{i=0}^{n}a_i\varphi_j(x) \tag{39}$$

要使（38）式达到极小值，则有：

$$\sum_{i=0}^{n}(\varphi_i,\varphi_k)a_i=(f,\varphi_k),\qquad k=0,1,\cdots n \tag{40}$$

用矩阵表示为：

$$\begin{bmatrix}(\varphi_0,\varphi_0) & (\varphi_0,\varphi_1) & \cdots(\varphi_0,\varphi_n)\\ (\varphi_1,\varphi_0) & (\varphi_1,\varphi_1) & (\varphi_1,\varphi_n)\\ \vdots & \vdots & \vdots\\ (\varphi_n,\varphi_0) & (\varphi_n,\varphi_1) & \cdots(\varphi_n,\varphi_n)\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}(f,\varphi_0)\\ (f,\varphi_1)\\ \vdots\\ (f,\varphi_n)\end{bmatrix} \tag{41}$$

其中，

$$(\varphi_i,\varphi_k)=\sum_{j=0}^{m}\rho(x_j)\varphi_i(x_j)\varphi_k(x_j),\qquad (f,\varphi_k)=\sum_{j=0}^{m}\rho(x_j)f(x_j)\varphi_k(x_j) \tag{42}$$

若系数矩阵为非奇异矩阵，能解出唯一一组解，其解就是所要拟合函数的系数，这样就可以确定黑龙江省各年份与就业率之间的函数关系，进而实现就业预测。

本文采用二次多项式进行拟合，其中 $\Phi=span\{1,x,x^2\}$，自变量 x 代表年份，因变量 f 代表每年的就业率，用矩阵表示即有：

$$\begin{bmatrix}(1,1) & (1,x) & (1,x^2)\\ (x,1) & (x,x) & (x,x^2)\\ (x^2,1) & (x^2,x) & (x^2,x^2)\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ a_2\end{bmatrix}=\begin{bmatrix}(1,f)\\ (x,f)\\ (x^2,f)\end{bmatrix} \tag{43}$$

为了便于计算，在此把自变量变为：$x-2003$，用 Matlab 求解，得出系数矩阵为：解得：

$a_0=0.7710$　　$a_1=-0.0031$　　$a_2=0.0017$　　（9 月份就业率拟合系数）

$a_0=0.7893$　　$a_1=0.0091$　　$a_2=0.0014$　　（12 月份就业率拟合系数）

所以，年份与大学生就业率的二次多项式模型为：

$f=0.7710-0.0031x+0.0017x^2$　　（9 月份就业率模型）　　(44)

$f=0.7893+0.0091x+0.0014x^2$　　（12 月份就业率模型）　　(45)

根据以上方程，可以反过来求出每年就业率的估计值，见下表：

由 $\delta_i=f_i-\hat{f}_i\quad i=1,2,\cdots,n$（$n$ 表示年份）

$$e=(\sum_{i=1}^{n}|\delta_i|)/n \tag{46}$$

$$\mu=(\sum_{i=1}^{n}\delta_i)/n \tag{47}$$

分别得出残差绝对值平均值和残差平均值为：

$e=0.53\%$　　$\mu=0.002\%$　　（9 月份）

$e=0.722\%$　　$\mu=0.002\%$　　（12 月份）

据此，可以得出预测结果与实际值的误差为：

下图为实际值与预测值比较图：

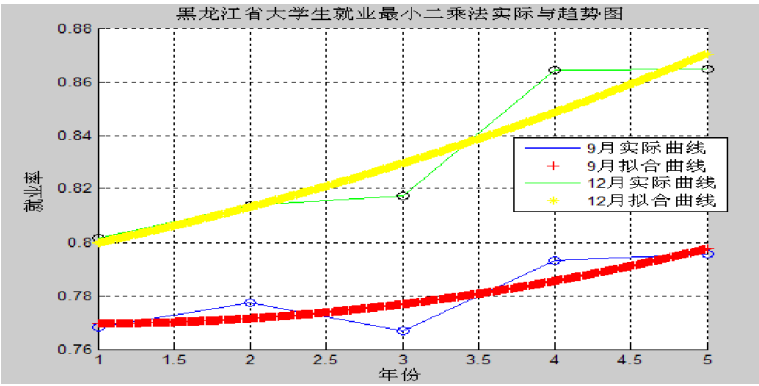


图 20 大学生就业率实际值与预测值对比图

由拟合曲线，预测 2009 年和 2010 黑龙江省大学生就业率如下表所示：

	2009 年	2010 年
9 月	0.8130	0.8953
12 月	0.8318	0.9229

可见。预测和实际基本吻合，大学生就业情况呈平稳递增趋势。

5.4 问题四（预测）

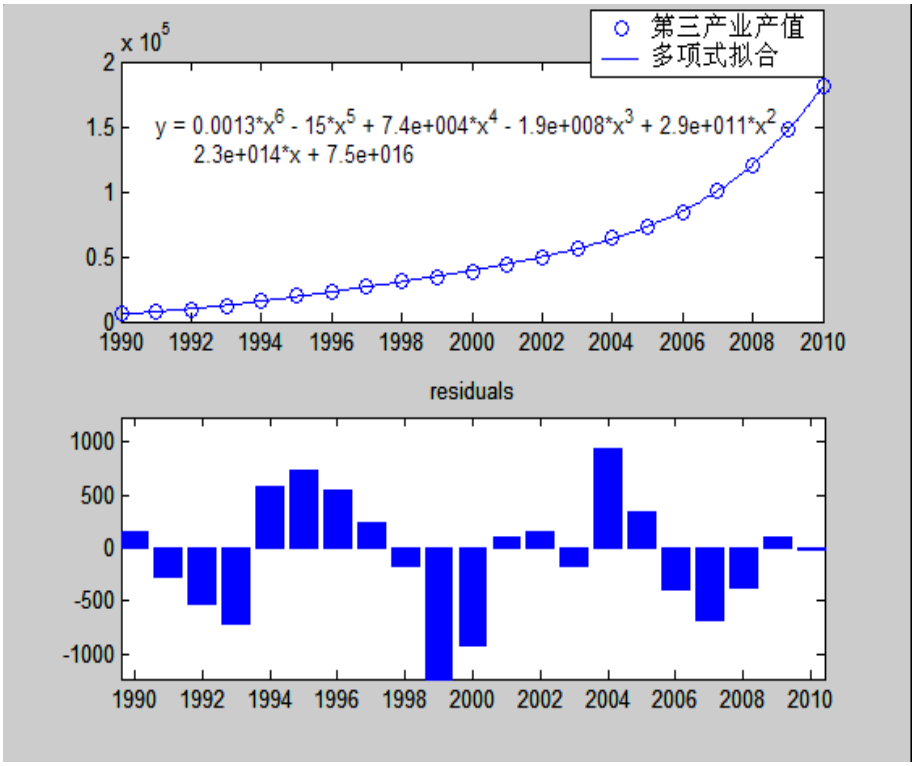
按照题目要求，本文利用第二问所建立的关于城镇就业人数的数学模型，对 2009 年及 2010 年上半年的我国就业前景进行仿真。其中，各项指标依然选为问题一中确定的 6 个指标：人均 GDP、第三产业产值、社会消费品零售总额、职工平均工资、政府财政支出以及总人口数。的数据来源于国家的有关决策和规划。

要对我国就业前景进行仿真，就必须具有 6 个影响就业的主要指标的数据。本文根据题目要求，对能从国家的有关决策和规划中查到的数据，就假定此数据为真实的，亦即假定国家的有关决策和规划能够达到。对于没有从国家相关政策和规划中查到的数据，本文根据历年的数据，通过多项式拟合进行预测。

例如：根据国家“十一五规划”，按现行汇率计算，2010 年我国,人均 GDP 将达到 2400 美元，为与之前的样本一致，按往年汇率折合成人民币进行预测。预计 2010 年总人口数控制在 13.6 亿内。2009 年中国财政预算为 7.6 万亿元人民币,较去年实际支出增长 22%[21]。

又例如：第三产业产值的数据很难从已有的规化中查到，有的能查到也不能查到具体的两年数据，因此要进行多项式拟合，来预测 2009 年和 2010 年的相应数值，将能查到的数据当作已知的数据来分析。具体过程如下：

用已有的数据在 Matlab 中进行曲线拟合，通过比对不同拟合方式下拟合曲线与实际值之间的符合程度，亦即残差大小，确定拟合曲线的相应参数，以第三产业产值为例，得到结果如下：



根据拟合得到的多项式形式，分别代入时间即可预测到 2009 年与 2010 年的相关指标的参数。

通过以上两种方法，本文得到了 2009 年和 2010 年的预测值。如下表所示：

表 5 2009 年和 2010 年预测值

时间	人均国内生产总值	社会消费品零售总额	职工平均工资	第三产业产值	人口总数	政府财政支出
2009	19200.0*	26847.0	133456.7	147302.8	133903.96	76000*
2010	24193.0	32193.0	168851.8	181447.2	135220.91	109842.90

**

*表示该量不是通过回归曲线预测得到

有了 2009 年和 2010 年的数据，就可以利用问题二中建立的因果关系理论和神经网络整合模型进行仿真，由此预测 2009 年和 2010 年的城镇就业人数。

由于仿真过程与问题二的处理类似，在这里将不再累述，有关过程的训练图详见附件。通过仿真，得出预测值分别为：

2009 年城镇就业人数为：31167 万人；

2010 年城镇就业人数为：31722 万人。

可见，中国整体的经济形势还是很乐观的。

5.5 问题五(对提高就业率的建议)

纵观整个建模过程及仿真结果，对提高我国城镇就业人口数提出如下几点建议：

(一) 尽快恢复经济增长

近年来，国际金融危机给我国带来了前所未有的困难和挑战。为应对国际金融危机冲击、保持经济平稳较快发展，中国及时调整宏观经济政策，果断实施积极的财政政策和适度宽松的货币政策，进一步扩大内需、促进经济增长，以此促进就业形势向更好的方向发展。

（二） 促进第三产业内部结构的升级

第三产业向来被认为是吸收劳动力能力最强的领域,无论是从相关性分析和Granger 因果检验，还是从反映城镇就业人数与第三产业的 VEC 定量模型，都可以看出第三产业与就业有着密不可分的关系。但是我国第三产业的发展速度不快,发展水平也不高,其吸收劳动力的潜力还没有释放出来。究其原因,可以从第三产业外部结构和内部结构两方面分析。

此外,根据国民经济分类标准第根据国民经济分类标准第二版 ISIC 关于三次产业的划分[11],与 1997 年各个不同收入水平的国家三次产业的就业结构(表 7)相比,我国第三产业就业比例不仅低于高收入水平和中等收入水平国家,而且低于许多低收入国家。第三产业与第二产业就业比是反映就业结构升级的重要指标,这个指标的数值我国也是很低的。

表 6 中国与世界其他国家产业结构对比

国家和地区		就业构成						第三产业就业与第二产业就业之比	
		第一产业		第二产业		第三产业			
		2000	2005	2000	2005	2000	2005	2000	2005
低收入国家	中国	50.0	44.8	22.5	23.8	27.5	31.4	1.2222	1.3193
	巴基斯坦	48.4	43.0	18.0	20.3	33.5	36.6	1.8611	1.8030
	印度尼西亚	45.1	44.0	17.5	18.0	37.3	38.0	2.1314	2.1111
	土耳其	36.0	29.5	24.0	24.7	40.0	45.8	1.6667	1.8543
中等收入国家	菲律宾	37.4	37.0	16.0	14.9	46.5	48.1	2.9063	3.2282
	泰国	48.8	42.6	19.0	20.2	32.2	37.1	1.6947	1.8366
高收入国家	美国	2.6	1.6	23.2	20.6	74.3	77.8	3.2026	3.7767
	日本	5.1	4.4	31.2	27.9	63.1	66.4	2.0224	2.3799

资料来源：世界银行数据库　　单位：%

基于上述分析及大量查阅有关资料，建议加快第三产业内部结构的升级。我国第三产业发展不充分的一个重要原因就是行业垄断和市场准入过严,如金融保险业、电信、交通、教育、卫生、广播电影电视业等,它们就业弹性较大,却存在很多进入壁垒。由于缺乏竞争,这些服务居高不下的价格不仅抑制了消费,反过来也影响了以上产业的发展规模,最终限制了对劳动力需求。所以,当前最重要的一点就是消除它们过高的进入门槛,吸引多方投资的加入。

（三） 学历不是就业的保证

中国社科院公布的一项研究报告显示，2007 年有 500 万年轻入毕业，有 145 万人没找到工作，只有 75％的应届毕业生找到工作。中国的很多学校都在夸大就业率。大学毕业生找工作越来越难，这是个不争的事实。失业率不断上升也反映了中国经济和教育体系中存在更广泛的问题。

随着经济的发展和科学技术的进步，各行各业对知识水平的要求也越来越高，因而学历被社会公认为敲开就业大门的金砖。毕业生应该认识到，学历不是就业的保证。学历是知识水平的象征，学历越高当然越有利于找工作，但也不能说学历高的就一定比学历低的找的工作就好。因为这还涉及到一个人的实力的问题。实力既要有扎实的专业功底，又要有较强的社会实践能力。

（四） 健全社会保障制度，扩大消费需求

去年，面对几场严重的突如其来的自然灾害，经历国际金融危机冲击的蔓延和持续，经历国际市场初级产品价格的大幅度下跌等复杂严峻的局面，不是靠某一个人所能解决的。社会主义的优越性由此体现：集中力量办大事。中央政府从 08 年 10 月开始实施了 40000 亿元的投资计划，确定了十大产业振兴计划，采取扩大国内消费需求的措施，提高对外开放水平以增加出口。

国家的这些宏观调控政策对缓解就业压力起了重要作用。政府的有关政策最好能不断地增加居民收入，不断地改善消费的环境，不断地改善、拓宽消费领域，不断地打消老百姓不愿意消费的顾虑，把社会保障制度更加健全起来，相信中国的消费会保持比较快的增长，随之而来的中国的就业率也无疑将增长。

(五) 坚持计划生育长期国策

中国人口基数庞大，每年新增劳动人口都在 2000 万上下。由于教育容量有限，每年需要就业的新增劳动人口则在 1000 万至 1600 万之间。另一方面，由于青年缺乏工作经验，在劳动力市场上竞争力不强。在劳动力市场供大于求的情况下，就业问题愈发突出，表现为失业率上升。

计划生育作为中国的一项基本国策，应该长期坚持。庞大的人口基数是短期内无法改变的事实，通过计划生育，控制人口数量，提高人口质量，有利于提高我国人民生活水平，有利于缓解就业压力。

(六) 加强国际合作

当今社会，是高速发展的社会，国际合作越来越密切，地球越来越像一个“地球村”。在金融危机的大背景下，更应该加强和其他各国经济、政治等各方面的合作与联系，共同抵抗金融危机，建设新的国际经济秩序。

加强国际合作，可以吸引外资，提供更多的就业机会，同时可以吸收和引进当代世界先进的科学技术成果，增加本国的经济实力，有利于解决金融危机中失业率等社会问题。

(七) 努力学习科学知识，提升职业技能

就个人而言，严峻的就业形势是短期内无法改变的事实，唯有努力提升自我，才是就业求职的硬道理。努力学习科学知识，确立知识、能力、素质协调发展，提高自身的文化修养，以适应飞速发展的现代社会，为自己在就业之路上多铺几块垫脚石。

抓住机会提升职业技能和适应社会的能力。比如，培养创新能力，专业技能，为人处事能力等。当个人素质达到一定高度时，就不用担心失业了。

6 模型的评价与改进

6.1 模型的优点

1） 数据收集较为真实。建立模型前，在数据的收集上做了大量工作。一方面，我们的数据大都是从官方网站（如：http://www.stats.gov.cn/tjsj/ndsj/《中国统计年鉴》）上整理得到的，这在最大程度上保证了数据的真实可靠性；另一方面，我

们搜集的多是近年的最新数据，这更加提高了数据的可靠性。

2) 所选数据较为详尽，能充分满足建模中对数据的需求。进行了细致的预处理，去粗取精，去伪存真，有利于在建模中很好地利用这些数据。例如，我们从官方网站搜查到了 1990 年到 2007 年的 15 个指标的详尽数据，并对数据进行了相关预处理。

3) 考虑的因素较为全面。本文综合考虑了人均 GDP、第三产业产值、社会零售商品总额、职工平均工资、政府财政支出、城镇就业人口比重、货物进出口总额、全社会固定资产投资额、政府财政支出、通货膨胀率、第一产业生产总值、第二产业生产总值、消费指数、发电量、钢材产量、能源生产总量、教育经费以及总人口数等 16 个有可能影响就业的指标。这样比主观地考虑少量几个因素更具说服力。

4) 建模过程思路清晰，过渡自然。从最简单的想法出发，层层递进，环环相扣，建立解决具体问题的最适合模型，并由此得出 6 方面指标对城镇就业人数的影响程度的定量权重。

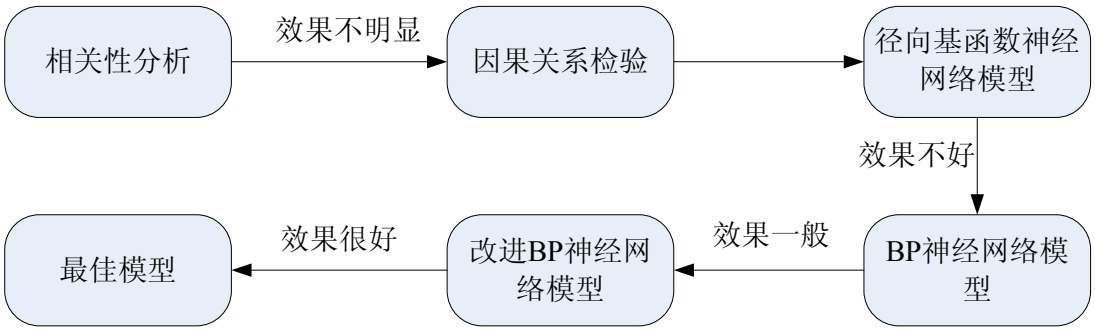


图 21 建模思路图（a）

5) 针对第三问，重点从经济社会指标、分行业、分地区、分就业人群四方面建立模型可加性分解模型、VEC 模型、区域划分、大学生就业率预测模型，其中既有定性分析，又有定量分析，模型详尽而精细。

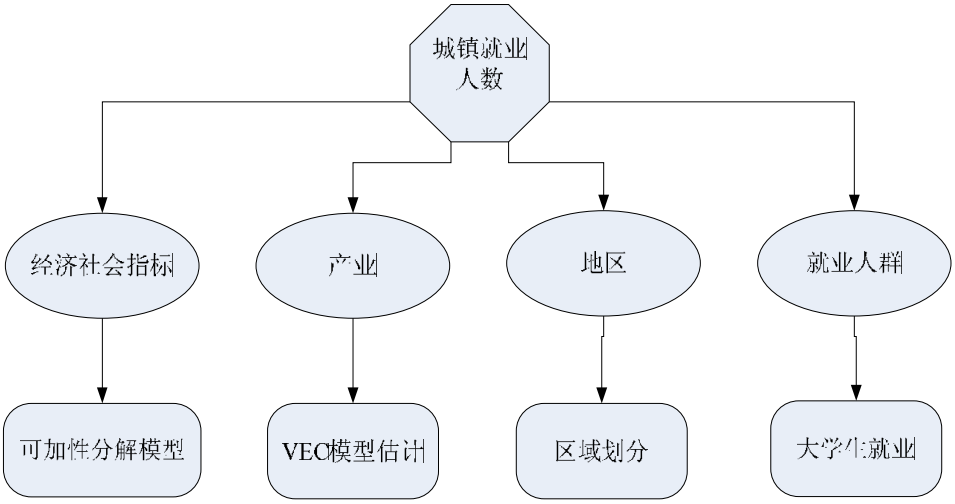


图 22 建模思路图（b）

6) 综合运用了 Matlab、Eviews、Spss、Excel 等软件进行数据处理，取长补短，提高了效率。

7） 考虑到了城镇就业人数的时滞性问题。

8） 我们还在建立模型和仿真（仿真过程充分考虑了国家政策调整所产生的影响）结果的基础之上，进行了简单的定性分析，为金融危机大背景下的国家宏观调控，企业发展和个人前途等提出了较为合理的建议。

6.2 模型缺点：

1） 鉴于城镇就业人数的复杂性，仅通过因果检验提取人均 GDP、第三产业产值、社会零售商品总额、职工平均工资、政府财政支出等影响就业的 6 个指标还远远不够，因为其他因素或多或少都会对城镇就业人数产生影响。

2) BP 神经网络结构的选择尚无一种统一而完整的理论指导，一般只能由经验选定。因此，应用中如何选择合适的网络结构是一个重要的问题。另外，BP 神经网络的重现性很差，训练失败的可能性较大，即使采用改进的 BP 神经网络，也不可避免地存在某些缺陷。

3) 虽然从网上搜到了大量数据，但并不能保证所有数据都真实可靠，这显然会给模型求解带来一定误差。

4) 短期内没有考虑由于突发事件所造成的影响，而是假定国家的规划恰好完成。但这并不意味着没有突发事件的影响。从而，模型应对特殊事件的能力还有待提高。

6.3 模型的改进方向

1）可以考虑的建立大型的、精细的、体现经济社会指标、行业、地区、就业人群等方面与城镇就业人数关系的模型。通过查阅有关资料，初步认为经济学中的 CEG 模型可能能够完成此重任。此问题有待进一步研究。

参考文献

[1] 李晓西，《宏观经济学（中国版）》，北京：中国人民大学出版社，2005 。

[2] 百度百科，产业结构，http://baike.baidu.com/view/61661.htm，2009-9-19 。

[3] 朱建平，《应用多元统计分析》，北京：科学出版社，2006。

[4] 王岩，隋思涟，王爱青，《数理统计与 MATLAB 工程数据分析》，北京：清华大学出 版 社，2006。

[5] 李静萍，谢邦昌，《多元统计分析方法与应用》，北京：中国人民大学出版社，2008。

[6] 林清泉，《计量经济学》，北京：中国人民大学出版社，2006。

[7] 孙敬水，《计量经济学教程》，北京：清华大学出版社，2005。

[8] 许涤龙，《Eviews 应用实验教程》，湖南：湖南大学出版社，2008。

[9] 董长虹，《Matlab 神经网络与应用》，第二版，北京：国防工业出版社，2007。

[10] 王正林，刘明，《精通 MATLAB7》，北京：电子工业出版社，2008。

[11] 飞思科技产品研发中心，《神经网络理论与 MATLAB7 实现》，北京：电子工业出版社， 2005。

[12] 韩力群，《人工神经网络教程》，北京：北京邮电大学出版社，2006

[13] 许长新，“关于劳动就业变化因素的经济模型”，《河海大学学报》，1997，25（3）：73-79。

[14] 崔亮，艾冰，“对产业结构与就业结构关系的探讨”，《财经问题研究》，2008（6）：112-116。

[15] 杨敏，韩春鲜，罗辉，“新疆旅游业与经济增长关系浅析——基于协整分析和 Granger 因果检验”，《干旱区资源与环境》，2007，23（4）：160-164。

[16] 王振波，朱传耿，“中国就业的空间模式及区域划分”，《地理学报》，2007，62（2）：191-198。

[17] 王振波，朱传耿，孙峰华，“山东省就业空间模式及机制研究”，《中国人口 资源与环境》，17（2）：131-135。

[18]郗杰英，杨守建，“青年就业的问题和对策：基于劳动力供求关系的分析”，《中国青年研究》 ，2005， (2): 25-30。

[19] 朱博，黑龙江——2008 年高校毕业生年终就业率 86.47%，http://www.chinanews.com.cn/edu/edu-qzjy/news/2009/06-11/1730026.shtml，2009-9-22。

[20] 李兴海，黑龙江省首次公布 2004 年高校毕业生年就业率，http://www.drcnet.com.cn/DRCnet.common.web/DocViewSummary.aspx?LeafID=103&DocID=1195688，2009-9-22。

[21] 国家人口计生委发展规划与信息司，关于“十一五”人口发展规划和计划生育事业发展规划中期评估情况的报告，http://www.chinapop.gov.cn/xwzx/zwyw/200904/t20090410_168701.html，2009-9-21。

7 附录

1. 用SPSS软件进行相关性分析得到的结果（各因素与城镇就业人口间的相关性）

Correlations[a]

		城镇就业人口	人均国内生产总值
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
人均国内生产总值	Pearson Correlation	.947**	
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=16

Correlations[a]

		城镇就业人口	城镇人口比重
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
城镇人口比重	Pearson Correlation	.993**	
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=15

Correlations[a]

		城镇就业人口	能源生产总量
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
能源生产总量	Pearson Correlation	.909**	
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed

a. Listwise N=15

Correlations

		城镇就业人口	通货膨胀率
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
通货膨胀率	Pearson Correlation	-.511*	
	Sig. (2-tailed)	.043	
	N	16	

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

		城镇就业人口	职工平均工资
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
职工平均工资	Pearson Correlation	.922**	
	Sig. (2-tailed)	.000	
	N	15	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	恩格尔系数
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
恩格尔系数	Pearson Correlation	-.953**	
	Sig. (2-tailed)	.000	
	N	15	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	固定资产总投资
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
固定资产总投资	Pearson Correlation	.919**	
	Sig. (2-tailed)	.000	
	N	13	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	平均劳动报酬
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
平均劳动报酬	Pearson Correlation	.990**	
	Sig. (2-tailed)	.000	
	N	10	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	第一产业生产总值
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
第一产业生产总值	Pearson Correlation	.968**	
	Sig. (2-tailed)	.000	
	N	16	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	第二产业生产总值
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
第二产业生产总值	Pearson Correlation	.930**	
	Sig. (2-tailed)	.000	
	N	16	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	发电量
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
发电量	Pearson Correlation	.950**	
	Sig. (2-tailed)	.000	
	N	13	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	钢材产量
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
钢材产量	Pearson Correlation	.917**	
	Sig. (2-tailed)	.000	
	N	13	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	零售总额
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
零售总额	Pearson Correlation	.983**	
	Sig. (2-tailed)	.000	
	N	13	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	教育经费
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
教育经费	Pearson Correlation	.977**	
	Sig. (2-tailed)	.000	
	N	13	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	第三产业生产总值
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
第三产业生产总值	Pearson Correlation	.934**	
	Sig. (2-tailed)	.000	
	N	16	

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		城镇就业人口	货物进出口总额
城镇就业人口	Pearson Correlation		
	Sig. (2-tailed)		
	N		
货物进出口总额	Pearson Correlation	.831**	
	Sig. (2-tailed)	.000	
	N	15	

**. Correlation is significant at the 0.01 level (2-tailed).

2. 用 Eviews 进行格兰杰因果检验得到的结果：

零售商品总额

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:04
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCDT does not Granger Cause LNWP	17	16.9266	0.0011
LNWP does not Granger Cause LNCDT		0.20971	0.6540

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:08
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCDT does not Granger Cause LNWP	15	5.68199	0.0221
LNWP does not Granger Cause LNCDT		13.3213	0.0018

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:07
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCDT does not Granger Cause LNWP	16	6.30419	0.0150
LNWP does not Granger Cause LNCDT		10.8327	0.0025

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:34
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCP does not Granger Cause LNWP	17	3.58971	0.0790
LNWP does not Granger Cause LNCP		1.18016	0.2957

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:41
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCP does not Granger Cause LNWP	16	2.45400	0.1315
LNWP does not Granger Cause LNCP		2.71769	0.1099

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:42
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCP does not Granger Cause LNWP	15	1.54790	0.2758
LNWP does not Granger Cause LNCP		1.99297	0.1937

发电量

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:43
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNEL does not Granger Cause LNWP	17	0.48462	0.4977
LNWP does not Granger Cause LNEL		0.47209	0.5032

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:44
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNEL does not Granger Cause LNWP	16	3.29568	0.0756
LNWP does not Granger Cause LNEL		7.47712	0.0089

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 13:45
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNCP does not Granger Cause LNWP	15	1.54790	0.2758
LNWP does not Granger Cause LNCP		1.99297	0.1937

恩格尔系数 1 阶：

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:15
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNEN does not Granger Cause LNWP	17	0.99482	0.3355
LNWP does not Granger Cause LNEN		0.32979	0.5749

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:15
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNEN does not Granger Cause LNWP	16	0.30661	0.7420
LNWP does not Granger Cause LNEN		1.52051	0.2612

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:16
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNEN does not Granger Cause LNWP	15	0.19615	0.8961
LNWP does not Granger Cause LNEN		1.38928	0.3147

能源生产总量：

1 阶

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:17
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNET does not Granger Cause LNWP	17	0.24624	0.6274
LNWP does not Granger Cause LNET		1.57280	0.2303

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:18
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNET does not Granger Cause LNWP	16	0.58466	0.5737
LNWP does not Granger Cause LNET		4.95397	0.0292

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:18
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNET does not Granger Cause LNWP	15	0.25219	0.8577
LNWP does not Granger Cause LNET		4.29779	0.0440

第一产值

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:20
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFG does not Granger Cause LNWP	17	21.9791	0.0003
LNWP does not Granger Cause LNFG		0.36344	0.5562

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:19
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFG does not Granger Cause LNWP	16	10.0422	0.0033
LNWP does not Granger Cause LNFG		2.59313	0.1195

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:20
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFG does not Granger Cause LNWP	15	8.62328	0.0069
LNWP does not Granger Cause LNFG		2.05632	0.1846

人均 GDP

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:21
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNGDP does not Granger Cause LNWP	17	13.3833	0.0026
LNWP does not Granger Cause LNGDP		0.82398	0.3794

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:21
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNGDP does not Granger Cause LNWP	16	8.07711	0.0069
LNWP does not Granger Cause LNGDP		15.6687	0.0006

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:22
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNGDP does not Granger Cause LNWP	15	13.4770	0.0017
LNWP does not Granger Cause LNGDP		4.50918	0.0393

进出口总额

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:23
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNIE does not Granger Cause LNWP	17	2.99368	0.1056
LNWP does not Granger Cause LNIE		3.08240	0.1010

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:24
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNIE does not Granger Cause LNWP	16	15.3409	0.0007
LNWP does not Granger Cause LNIE		2.67941	0.1127

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:24
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNIE does not Granger Cause LNWP	15	8.45944	0.0073
LNWP does not Granger Cause LNIE		3.59004	0.0658

二产产值

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:25
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNSG does not Granger Cause LNWP	17	11.6378	0.0042
LNWP does not Granger Cause LNSG		1.08731	0.3147

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:25
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNSG does not Granger Cause LNWP	16	6.25167	0.0154
LNWP does not Granger Cause LNSG		22.2629	0.0001

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:26
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNSG does not Granger Cause LNWP	15	4.26493	0.0448
LNWP does not Granger Cause LNSG		5.88961	0.0201

三产产值

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:28
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNTG does not Granger Cause LNWP	17	10.0343	0.0068
LNWP does not Granger Cause LNTG		1.92784	0.1867

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:28
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNTG does not Granger Cause LNWP	16	6.71755	0.0124
LNWP does not Granger Cause LNTG		4.14597	0.0455

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:29
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNTG does not Granger Cause LNWP	15	12.3023	0.0023
LNWP does not Granger Cause LNTG		3.24018	0.0816

总人口

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:35
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNTP does not Granger Cause LNWP	17	7.66580	0.0151
LNWP does not Granger Cause LNTP		6.98324	0.0193

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:36
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNTP does not Granger Cause LNWP	16	3.24124	0.0782
LNWP does not Granger Cause LNTP		1.04243	0.3850

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:36
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNTP does not Granger Cause LNWP	15	3.59100	0.0658
LNWP does not Granger Cause LNTP		1.39240	0.3139

固定资产投入

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:42
Sample: 1994 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNINPUT does not Granger Cause LNWP	13	0.62062	0.4491
LNWP does not Granger Cause LNINPUT		0.01780	0.8965

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:43
Sample: 1994 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNINPUT does not Granger Cause LNWP	12	9.02704	0.0115
LNWP does not Granger Cause LNINPUT		1.41017	0.3058

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:43
Sample: 1994 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNINPUT does not Granger Cause LNWP	11	0.72426	0.5882
LNWP does not Granger Cause LNINPUT		2.53214	0.1955

职工平均 工资

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:46
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNAS does not Granger Cause LNWP	17	11.9864	0.0038
LNWP does not Granger Cause LNAS		0.63317	0.4395

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:47
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNAS does not Granger Cause LNWP	16	26.1199	7.E-05
LNWP does not Granger Cause LNAS		4.54818	0.0364

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 14:48
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNAS does not Granger Cause LNWP	15	22.2702	0.0003
LNWP does not Granger Cause LNAS		4.16890	0.0472

财政支出:

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 15:30
Sample: 1990 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFO does not Granger Cause LNWP	17	9.46826	0.0082
LNWP does not Granger Cause LNFO		0.02908	0.8670

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 15:31
Sample: 1990 2007
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFO does not Granger Cause LNWP	16	10.3405	0.0030
LNWP does not Granger Cause LNFO		2.09627	0.1693

Pairwise Granger Causality Tests
Date: 09/21/09 Time: 15:32
Sample: 1990 2007
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
LNFO does not Granger Cause LNWP	15	5.12195	0.0288
LNWP does not Granger Cause LNFO		1.52235	0.2816

ADF 单位根检验结果

Null Hypothesis: D(CDT,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.355077	0.3840
Test critical values:	1% level	-4.728363	
	5% level	-3.759743	
	10% level	-3.324976	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 15

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CDT,3)
Method: Least Squares
Date: 09/20/09 Time: 21:35
Sample (adjusted): 1993 2007
Included observations: 15 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(CDT(-1),2)	-0.753677	0.320022	-2.355077	0.0364
C	-514.2552	851.7942	-0.603732	0.5573
@TREND(1990)	112.9073	79.97206	1.411835	0.1834

R-squared	0.348793	Mean dependent var	206.9333
Adjusted R-squared	0.240259	S.D. dependent var	1501.094
S.E. of regression	1308.400	Akaike info criterion	17.36785
Sum squared resid	20542935	Schwarz criterion	17.50946
Log likelihood	-127.2589	Hannan-Quinn criter.	17.36635
F-statistic	3.213663	Durbin-Watson stat	1.881579
Prob(F-statistic)	0.076263		

城市人口比重:

Null Hypothesis: D(CP,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.411903	0.0015
Test critical values:	1% level	-4.992279	
	5% level	-3.875302	
	10% level	-3.388330	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 12

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CP,3)
Method: Least Squares
Date: 09/20/09 Time: 21:44
Sample (adjusted): 1996 2007
Included observations: 12 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(CP(-1),2)	-3.466693	0.540665	-6.411903	0.0007
D(CP(-1),3)	1.579049	0.413485	3.818879	0.0088
D(CP(-2),3)	0.956952	0.295913	3.233898	0.0178
D(CP(-3),3)	0.378370	0.188838	2.003673	0.0920
C	1.548305	0.264491	5.853911	0.0011
@TREND(1990)	-0.119804	0.020773	-5.767396	0.0012

R-squared	0.928667	Mean dependent var	0.010000
Adjusted R-squared	0.869222	S.D. dependent var	0.432162
S.E. of regression	0.156284	Akaike info criterion	-0.567437
Sum squared resid	0.146547	Schwarz criterion	-0.324984
Log likelihood	9.404623	Hannan-Quinn criter.	-0.657202
F-statistic	15.62243	Durbin-Watson stat	1.596039
Prob(F-statistic)	0.002194		

发电量

Null Hypothesis: D(EL,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.618348	0.7235
Test critical values:	1% level	-4.992279	
	5% level	-3.875302	
	10% level	-3.388330	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 12

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EL,3)
Method: Least Squares
Date: 09/20/09 Time: 21:56
Sample (adjusted): 1996 2007
Included observations: 12 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(EL(-1),2)	-1.005290	0.621183	-1.618348	0.1567
D(EL(-1),3)	0.761636	0.482721	1.577798	0.1657
D(EL(-2),3)	-0.295292	0.369245	-0.799718	0.4544
D(EL(-3),3)	0.819229	0.348801	2.348703	0.0572
C	-239.2135	377.5725	-0.633557	0.5497
@TREND(1990)	40.64433	41.96429	0.968546	0.3702

R-squared	0.836753	Mean dependent var	50.02583
Adjusted R-squared	0.700714	S.D. dependent var	378.4660
S.E. of regression	207.0476	Akaike info criterion	13.81063
Sum squared resid	257212.3	Schwarz criterion	14.05308
Log likelihood	-76.86377	Hannan-Quinn criter.	13.72086
F-statistic	6.150823	Durbin-Watson stat	2.607510
Prob(F-statistic)	0.023482		

恩格尔系数：

Null Hypothesis: D(EN,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.177584	0.0251
Test critical values:	1% level	-4.728363	
	5% level	-3.759743	
	10% level	-3.324976	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 15

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EN,3)
Method: Least Squares
Date: 09/20/09 Time: 22:01
Sample (adjusted): 1993 2007
Included observations: 15 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(EN(-1),2)	-1.213041	0.290369	-4.177584	0.0013
C	-0.520589	0.841245	-0.618832	0.5476
@TREND(1990)	0.060015	0.077202	0.777378	0.4520

R-squared	0.594065	Mean dependent var	0.120000
Adjusted R-squared	0.526409	S.D. dependent var	1.866318
S.E. of regression	1.284362	Akaike info criterion	3.515258
Sum squared resid	19.79504	Schwarz criterion	3.656868
Log likelihood	-23.36444	Hannan-Quinn criter.	3.513750
F-statistic	8.780675	Durbin-Watson stat	2.010616
Prob(F-statistic)	0.004474		

能源总量

Null Hypothesis: D(ET,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.483675	0.0780
Test critical values:	1% level	-4.728363	
	5% level	-3.759743	
	10% level	-3.324976	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 15

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ET,3)
Method: Least Squares
Date: 09/20/09 Time: 22:04
Sample (adjusted): 1993 2007
Included observations: 15 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(ET(-1),2)	-1.008445	0.289477	-3.483675	0.0045
C	667.0064	4225.721	0.157844	0.8772
@TREND(1990)	13.94878	387.8440	0.035965	0.9719

R-squared	0.502926	Mean dependent var	-152.0667
Adjusted R-squared	0.420080	S.D. dependent var	8518.073
S.E. of regression	6486.724	Akaike info criterion	20.56976
Sum squared resid	5.05E+08	Schwarz criterion	20.71137
Log likelihood	-151.2732	Hannan-Quinn criter.	20.56825
F-statistic	6.070628	Durbin-Watson stat	1.995157
Prob(F-statistic)	0.015084		

一产产值

Null Hypothesis: FG has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.357172	0.0203
Test critical values:	1% level	-4.800080	
	5% level	-3.791172	
	10% level	-3.342253	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 14

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(FG)
Method: Least Squares
Date: 09/20/09 Time: 22:06
Sample (adjusted): 1994 2007
Included observations: 14 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
FG(-1)	-1.598953	0.366970	-4.357172	0.0024
D(FG(-1))	0.812700	0.248039	3.276496	0.0112
D(FG(-2))	0.636336	0.283606	2.243732	0.0551
D(FG(-3))	1.415363	0.355617	3.980025	0.0041
C	5132.661	1319.237	3.890627	0.0046
@TREND(1990)	1707.761	382.0302	4.470226	0.0021

R-squared	0.743450	Mean dependent var	1509.371
Adjusted R-squared	0.583107	S.D. dependent var	1349.794
S.E. of regression	871.5256	Akaike info criterion	16.67589
Sum squared resid	6076455.	Schwarz criterion	16.94978
Log likelihood	-110.7313	Hannan-Quinn criter.	16.65054
F-statistic	4.636606	Durbin-Watson stat	1.836108
Prob(F-statistic)	0.027724		

人均 GDP

Null Hypothesis: D(GDP,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.518012	0.3158
Test critical values:	1% level	-4.800080	
	5% level	-3.791172	
	10% level	-3.342253	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 14

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GDP,3)
Method: Least Squares
Date: 09/20/09 Time: 22:11
Sample (adjusted): 1994 2007
Included observations: 14 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(GDP(-1),2)	-0.961645	0.381907	-2.518012	0.0305
D(GDP(-1),3)	0.213031	0.304406	0.699825	0.5000
C	-256.8597	208.5059	-1.231906	0.2462
@TREND(1990)	37.85658	19.44063	1.947292	0.0801

R-squared	0.469240	Mean dependent var	27.71429
Adjusted R-squared	0.310013	S.D. dependent var	333.3481
S.E. of regression	276.8973	Akaike info criterion	14.32013
Sum squared resid	766720.9	Schwarz criterion	14.50271
Log likelihood	-96.24089	Hannan-Quinn criter.	14.30322
F-statistic	2.946974	Durbin-Watson stat	2.047209
Prob(F-statistic)	0.084970		

进出口总额

Null Hypothesis: D(IE,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.689249	0.0588
Test critical values:	1% level	-4.800080	
	5% level	-3.791172	
	10% level	-3.342253	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 14

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(IE,3)
Method: Least Squares
Date: 09/20/09 Time: 22:15
Sample (adjusted): 1994 2007
Included observations: 14 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(IE(-1),2)	-1.664926	0.451291	-3.689249	0.0042
D(IE(-1),3)	0.396095	0.291991	1.356532	0.2048
C	-1598.723	3973.796	-0.402316	0.6959
@TREND(1990)	406.5679	369.7347	1.099621	0.2973

R-squared	0.657230	Mean dependent var	104.4000
Adjusted R-squared	0.554399	S.D. dependent var	7891.586
S.E. of regression	5267.901	Akaike info criterion	20.21161
Sum squared resid	2.78E+08	Schwarz criterion	20.39420
Log likelihood	-137.4813	Hannan-Quinn criter.	20.19471
F-statistic	6.391351	Durbin-Watson stat	1.823987
Prob(F-statistic)	0.010820		

二产产值

Null Hypothesis: D(SG,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.978928	0.1711
Test critical values:	1% level	-4.800080	
	5% level	-3.791172	
	10% level	-3.342253	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 14

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SG,3)
Method: Least Squares
Date: 09/20/09 Time: 22:18
Sample (adjusted): 1994 2007
Included observations: 14 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(SG(-1),2)	-0.853550	0.286529	-2.978928	0.0138
D(SG(-1),3)	0.297239	0.262374	1.132883	0.2837
C	-1626.357	1126.875	-1.443246	0.1795
@TREND(1990)	231.0611	109.1762	2.116404	0.0604

R-squared	0.495562	Mean dependent var	18.87857
Adjusted R-squared	0.344231	S.D. dependent var	1812.811
S.E. of regression	1468.007	Akaike info criterion	17.65616
Sum squared resid	21550447	Schwarz criterion	17.83874
Log likelihood	-119.5931	Hannan-Quinn criter.	17.63925
F-statistic	3.274682	Durbin-Watson stat	2.006745
Prob(F-statistic)	0.067195		

三产产值

Null Hypothesis: D(TG,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.325270	0.8325
Test critical values:	1% level	-4.886426	
	5% level	-3.828975	
	10% level	-3.362984	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 13

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TG,3)
Method: Least Squares
Date: 09/20/09 Time: 22:21
Sample (adjusted): 1995 2007
Included observations: 13 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(TG(-1),2)	-0.724966	0.547033	-1.325270	0.2217
D(TG(-1),3)	-0.350148	0.427142	-0.819748	0.4361
D(TG(-2),3)	-0.597698	0.281916	-2.120130	0.0668
C	-2097.390	740.3875	-2.832827	0.0221
@TREND(1990)	254.8076	75.85962	3.358936	0.0099

R-squared	0.782386	Mean dependent var	179.8308
Adjusted R-squared	0.673579	S.D. dependent var	1458.063
S.E. of regression	833.0390	Akaike info criterion	16.57176
Sum squared resid	5551632	Schwarz criterion	16.78905
Log likelihood	-102.7164	Hannan-Quinn criter.	16.52710
F-statistic	7.190590	Durbin-Watson stat	2.332921
Prob(F-statistic)	0.009261		

人口

Null Hypothesis: D(TP,2) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.076876	0.0080
Test critical values:	1% level	-3.959148	
	5% level	-3.081002	
	10% level	-2.681330	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 15

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TP,3)
Method: Least Squares
Date: 09/20/09 Time: 22:27
Sample (adjusted): 1993 2007
Included observations: 15 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(TP(-1),2)	-0.967828	0.237395	-4.076876	0.0013
C	-42.75513	16.17887	-2.642653	0.0203

R-squared	0.561121	Mean dependent var	8.733333
Adjusted R-squared	0.527361	S.D. dependent var	56.96674
S.E. of regression	39.16392	Akaike info criterion	10.29695
Sum squared resid	19939.56	Schwarz criterion	10.39136
Log likelihood	-75.22716	Hannan-Quinn criter.	10.29595
F-statistic	16.62092	Durbin-Watson stat	1.385986
Prob(F-statistic)	0.001309		

就业人口

Null Hypothesis: D(WP,2) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.998715	0.0093
Test critical values:	1% level	-3.959148	
	5% level	-3.081002	
	10% level	-2.681330	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 15

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(WP,3)
Method: Least Squares
Date: 09/20/09 Time: 22:29
Sample (adjusted): 1993 2007
Included observations: 15 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(WP(-1),2)	-1.094811	0.273791	-3.998715	0.0015
C	46.44132	36.96822	1.256250	0.2311

R-squared	0.551565	Mean dependent var	5.933333
Adjusted R-squared	0.517070	S.D. dependent var	198.1444
S.E. of regression	137.6968	Akaike info criterion	12.81155
Sum squared resid	246485.3	Schwarz criterion	12.90596
Log likelihood	-94.08663	Hannan-Quinn criter.	12.81055
F-statistic	15.98972	Durbin-Watson stat	2.047692
Prob(F-statistic)	0.001516		

3. 神经网络训练的样本数据：

年份	就业人口	人均GDP	社会消费品零售	职工平均工	第三产业产值	人口总数	政府财政支
1990	17041	1644	8300.1	2140	5888.4	114333	3083.59
1991	17465	1893	9415.6	2340	7337.1	115823	3386.62
1992	17861	2311	10993.7	2711	9357.4	117171	3742.20
1993	18262	2998	14270.4	3371	11915.7	118517	4642.30
1994	18653	4044	18622.9	4538	16179.8	119850	5792.62
1995	19040	5046	23613.8	5500	19978.5	121121	6823.72
1996	19922	5846	28360.2	6210	23326.2	122389	7937.55
1997	20781	6420	31252.9	6470	26988.1	123626	9233.56
1998	21616	6796	33378.1	7479	30580.5	124761	10798.18
1999	22412	7159	35647.9	8346	33873.4	125786	13187.67
2000	23151	7858	39105.7	9371	38714.0	126743	15886.50
2001	23940	8622	43055.4	10870	44361.6	127627	18902.58
2002	24780	9398	48135.9	12422	49898.9	128453	22053.15
2003	25639	10542	52516.3	14040	56004.7	129227	24649.95
2004	26476	12336	59501.0	16024	64561.3	129988	28486.89
2005	27331	14053	67176.6	18364	73432.9	130756	33930.28
2006	28310	16165	76410.0	21001	84721.4	131448	40422.73
2007	29350	18934	89210.0	24932	100053.5	132129	49781.35

4. 神经网络的程序：

BP 神经网络：

```
clc;
p = [0 0   0   0   0   0;...
0.020046711   0.018946439   0.012327416   0.02144776 0.09072642
    0.009823744;...
0.053758364   0.045750002   0.035194773   0.051358097   0.17280643
    0.021351075;...
```

```
0.109143713   0.101403786   0.075875247   0.089234699   0.254764659
    0.050530867;...
0.193408342   0.175329716   0.14780572 0.15236398 0.335931316
    0.087822389;...
0.27413409 0.260098681   0.207100592   0.208603953   0.413322779
    0.12124899;...
0.338616049   0.340714886   0.250862919   0.258168094   0.490531572
    0.157357564;...
0.384896485   0.389846543   0.26688856 0.312382834   0.565852767
    0.199372121;...
0.415184976   0.425942439   0.329080375   0.365567351   0.634963161
    0.250094581;...
0.444395324   0.464494323   0.382519724   0.414320125   0.697375632
    0.327557997;...
0.500739471   0.523224037   0.445697732   0.485984186   0.755647567
    0.415049718;...
0.562310115   0.590308527   0.538091716   0.56959812 0.809474517
    0.512826174;...
0.624873435   0.676599322   0.633752465   0.651578088   0.859769835
    0.614962578;...
0.717057872   0.750999125   0.733481262   0.741975227   0.906898861
    0.699146651;...
0.861598649   0.869632196   0.855769231   0.868655754   0.953236315
    0.823534065;...
1   1   1   1   1   1]';
t = [0 0.041205053 0.079689018 0.118658892 0.156656948 0.194266278 0.279980564
0.36345967  0.444606414  0.521963071  0.593780369  0.670456754  0.752089407
0.835568513 0.916909621 1];

net = newff(minmax(p),[5 1],{'tansig','purelin'},'trainbfg');
net.trainParam.show = 5;
net.trainParam.epchos = 500;
net.trainParam.goal = 0.0001;
net.trainParam.lr=2;
inputWeights=net.IW{1,1}
[net,tr] = train(net,p,t);
MSE=mse(t-sim(net,p))
%info = trainbfg(code)
p1 = [1.170199049 1.156826578   1.162536982   1.167127499   1.042136029
    1.210474771]';
p2=[1.393343541   1.374230805   1.404832347   1.394120264   1.083602265
    1.513866155 ]';


goal=[p1,p2];
future = [1.414779595   2.125747285   2.443478797   1.697465349
    1.191680083   2.363832554;
1.817148844   2.726922626   3.397312623   2.093649306   1.271869086
    2.812597073]';
a = sim(net,future)
    tt=1:1:size(p,2);
    figure()
    plot(tt,t,'+',tt,sim(net,p))
    figure()
    plot(tt,t-sim(net,p))
  %  figure()
   % plotpc(net.iw{1},net.b{1});
```

径向基函数：

```
p = [0 0   0   0   0   0;...
0.020046711   0.018946439   0.012327416   0.02144776 0.09072642
    0.009823744;...
```

```
0.053758364   0.045750002   0.035194773   0.051358097   0.17280643
   0.021351075;...
0.109143713   0.101403786   0.075875247   0.089234699   0.254764659
   0.050530867;...
0.193408342   0.175329716   0.14780572 0.15236398 0.335931316
   0.087822389;...
0.27413409 0.260098681   0.207100592   0.208603953   0.413322779
   0.12124899;...
0.338616049   0.340714886   0.250862919   0.258168094   0.490531572
   0.157357564;...
0.384896485   0.389846543   0.26688856 0.312382834   0.565852767
   0.199372121;...
0.415184976   0.425942439   0.329080375   0.365567351   0.634963161
   0.250094581;...
0.444395324   0.464494323   0.382519724   0.414320125   0.697375632
   0.327557997;...
0.500739471   0.523224037   0.445697732   0.485984186   0.755647567
   0.415049718;...
0.562310115   0.590308527   0.538091716   0.56959812 0.809474517
   0.512826174;...
0.624873435   0.676599322   0.633752465   0.651578088   0.859769835
   0.614962578;...
0.717057872   0.750999125   0.733481262   0.741975227   0.906898861
   0.699146651;...
0.861598649   0.869632196   0.855769231   0.868655754   0.953236315
   0.823534065;...
1   1   1   1   1   1]';
t = [0 0.041205053 0.079689018 0.118658892 0.156656948 0.194266278 0.279980564
0.36345967  0.444606414  0.521963071  0.593780369  0.670456754  0.752089407
0.835568513 0.916909621 1];
pp=[1644   8300.1 2140   5888.4 114333 3083.59;...
1893   9415.6 2340   7337.1 115823 3386.62;...
2311   10993.7    2711   9357.4 117171 3742.20;...
2998   14270.4    3371   11915.7    118517 4642.30;...
4044   18622.9    4538   16179.8    119850 5792.62;...
5046   23613.8    5500   19978.5    121121 6823.72;...
5846   28360.2    6210   23326.2    122389 7937.55;...
6420   31252.9    6470   26988.1    123626 9233.56;...
6796   33378.1    7479   30580.5    124761 10798.18;...
7159   35647.9    8346   33873.4    125786 13187.67;...
7858   39105.7    9371   38714.0    126743 15886.50;...
8622   43055.4    10870  44361.6    127627 18902.58;...
9398   48135.9    12422  49898.9    128453 22053.15;...
10542  52516.3    14040  56004.7    129227 24649.95;...
12336  59501.0    16024  64561.3    129988 28486.89;...
14053  67176.6    18364  73432.9    130756 33930.28]';
tt=[17041 17465 17861 18262 18653 19040 19922 20781 21616 22412 23151 23940 24780
25639 26476 27331];
SPREAD=1000;
net = newrbe(p,t,SPREAD);
%net.trainParam.show = 5;
%net.trainParam.epchos = 500;
%net.trainParam.goal = 0.0001;
%net.trainParam.lr=2;
%[net,tr] = train(net,p,t);
inputWeights=net.IW{1,1}
%MSE = mse()

p1 = [14053   67176.6    18364  73432.9    130756 33930.28]';
p2=[16165  76410.0    21001  84721.4    131448 40422.73]';
goal=[p1,p2];
```

```
future = [1.414779595   2.125747285   2.443478797   1.697465349
   1.191680083   2.363832554;
1.817148844   2.726922626   3.397312623   2.093649306   1.271869086
   2.812597073]';
a = sim(net,future)
```

5. 主成分分析结果图：

图 方差解释及因子提取结果图

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.701	61.585	61.585	11.701	61.585	61.585
2	3.643	19.175	80.760	3.643	19.175	80.760
3	1.433	7.543	88.302	1.433	7.543	88.302
4	.692	3.643	91.945			
5	.431	2.266	94.212			
6	.334	1.758	95.969			
7	.236	1.240	97.209			
8	.181	.951	98.160			
9	.104	.546	98.706			
10	.060	.314	99.020			
11	.047	.247	99.267			
12	.044	.230	99.498			
13	.028	.149	99.647			
14	.023	.122	99.769			
15	.017	.092	99.861			
16	.013	.069	99.930			
17	.008	.040	99.970			
18	.004	.019	99.989			
19	.002	.011	100.000			

Extraction Method: Principal Component Analysis.

图 因子载荷矩阵

Component Matrix [a]

	Component		
	1	2	3
农林牧渔业	-.006	.172	.836
采掘业	.293	.584	.503
制造业	.786	.217	-.286
电力水燃气供应业	.775	.576	.139
建筑业	.672	.386	-.296
交通运输仓储邮政业	.947	-.069	.133
信息传输计算机软件业	.777	-.584	.046
批发零售业	.948	.041	.028
住宿餐饮业	.884	-.382	-.143
金融业	.949	.165	-.117
房地产行业	.825	-.496	-.089
租赁商务服务业	.734	-.653	-.021
科技服务地质业	.770	-.530	.160
水利环境公共管理	.884	.335	.148
居民服务业和其他服务业	.634	-.518	.386
教育业	.801	.544	-.108
卫生社保福利业	.887	.411	-.115
文体娱乐业	.904	-.332	.061
公共管理和社会组织业	.788	.561	-.068

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

6. 聚类分析结果图：

表 最终类中心数据

	Cluster		
	1	2	3
VAR00001	6.00	29.30	49.01
VAR00002	31.40	36.30	19.39
VAR00003	62.60	34.40	31.60

图 地区所属类分布结果

Cluster Membership

Case Number	Region	Cluster	Distance
1	北京	1	11.951
2	天津	2	13.840
3	河北	2	15.474
4	山西	3	10.715
5	内蒙古	3	4.471
6	辽宁	2	13.520
7	吉林	3	2.986
8	黑龙江	3	3.350
9	上海	1	11.951
10	江苏	2	9.940
11	浙江	2	13.815
12	安徽	3	7.734
13	福建	2	3.876
14	江西	3	10.146
15	山东	2	9.733
16	河南	3	10.299
17	湖北	3	12.855
18	湖南	3	2.549
19	广东	2	3.891
20	广西	3	9.236
21	海南	3	10.640
22	重庆	3	12.100
23	四川	3	3.325
24	贵州	3	9.658
25	云南	3	19.334
26	西藏	3	12.288
27	陕西	3	.660
28	甘肃	3	7.356
29	青海	3	5.917
30	宁夏	3	4.687
31	新疆	3	6.966